COMMENTATIONES SCIENTIARUM SOCIALIUM 66

Per Mickwitz

ENVIRONMENTAL POLICY EVALUATION: CONCEPTS AND PRACTICE

Published by The Finnish Society of Sciences and Letters
Cover design by Iiro Törmä / Törmä & Baer

ISSN 0355-256X
ISBN: 951-653-334-5

ISBN 978-951-44-9424-6(pdf)

Printed by Gummerus Oy, Saarijärvi 2006

Dedicated to the memory of Gösta

CONTENTS

The author and the publisher wish to thank the following who have kindly given permission for the use of copyright material in the appendices.

Elsevier Science B.V. for the article: Mickwitz Per 2003. Is it as bad as it sounds or as good as it looks? Experiences of Finnish water discharge limits. *Ecological Economics* 45(2): 237-255.

SAGE Publications Ltd. for the articles: Mickwitz Per 2003. A Framework for Evaluating Environmental Policy Instruments: Context and Key Concepts. *Evaluation* 9(4): 415-436. and Mickwitz, Per and Paula Kivimaa 2005. Evaluating Policy Integration – The Case of Policies for Environmentally Friendlier Technological Innovations. *Evaluation*, accepted on September 9th 2005, Volume and issue not yet assigned.

The Finnish Evaluation Society for the Article: Mickwitz Per 2002. Effectiveness Evaluation of Environmental Policy: the Role of Intervention Theories. *Hallinnon Tutkimus* (Administrative Studies) 21 (4): 77-87.

ORIGINAL PUBLICATIONS AND MANUSCRIPTS

This thesis is based on the following five articles and manuscripts, which are published as appendices in this book and will be referred to in the text by the Roman number of the appendix.

Mickwitz Per. 2003. "A Framework for Evaluating Environmental Policy Instruments: Context and Key Concepts". *Evaluation* 9(4): 415-436.

Mickwitz Per 2003. "Is it as bad as it sounds or as good as it looks? Experiences of Finnish water discharge limits". *Ecological Economics* 45(2): 237-255.

Mickwitz Per 2002. "Effectiveness Evaluation of Environmental Policy: the Role of Intervention Theories". *Hallinnon Tutkimus* (Administrative Studies) 21 (4): 77-87.

Mickwitz Per and Mikael Hildén 2004. "Multi-criteria Evaluations for Environmental Policy Development – The Case of Finnish Permit Systems". Manuscript.

Mickwitz Per and Paula Kivimaa 2005. "Evaluating Policy Integration – The Case of Policies for Environmentally Friendlier Technological Innovations". *Evaluation*, accepted on September 9th 2005.

AUTHOR'S CONTRIBUTION

The author of this thesis is fully responsible for Articles I, II and III.

Manuscript IV was written together with Professor Mikael Hildén. The author of this thesis took the initiative for the paper and was responsible for the work on multiple criteria and linking the analysis to the general evaluation debate. The paper was jointly written.

Manuscript V was written with researcher Paula Kivimaa. The author of this thesis initiated the paper and was responsible for the sections on: key concepts; evaluation of policy integration; and integration of innovation aims into environmental policies, while Paula Kivimaa was responsible for examining the integration of environmental aims into technology policies. Writing the manuscript and drawing the conclusions was as a collaborative effort.

once we were people
who used to dream about the future
once we were people with stars in our eyes

Erin Moran

1. INTRODUCTION

The task of this thesis is to examine concepts, which can be used when evaluating environmental policies. But, since evaluation is of necessity based on empirics it would not be meaningful to explore it solely conceptually. The usefulness of the approaches, methods and criteria will, in the end, be determined through their actual use, that is, through the practice of evaluation. The practice of evaluation utilising the proposed concepts makes special reference to experiences from an evaluation of Finnish environmental permits. The main aims of this thesis can be summarised in two questions: *"Should environmental policies be evaluated?"* and if so *"How could they be evaluated?"*.

This thesis is not about the concepts and practice of all kinds of evaluation. It is focused on public policy evaluation and limited to environmental policies. This means that the approaches, methods and criteria should be appropriate when the features of public policies and environmental issues are taken into account. Features more relevant, for example, to product or employee evaluations are not central, even if dealing with the environmental aspects of the products or employees in the environmental administration.

Evaluation, following Scriven (1991, 139) *"refers to the process of determining the* ***merit***, ***worth***, *or* ***value*** [emphasis in the original] *of something, or the product of that process."* Environmental policy evaluation will, in this thesis, be considered as "careful assessment of the merit, worth and value of administration, output and outcome of environmental policies, which is intended to play a role in future, practical action situations." (as discussed in Article I, based on Vedung 1997, 3) Environmental policies are considered as: *"the efforts by which public authorities wield their power in attempting to affect society in terms of values and beliefs, action and organisation in such a way as to improve, or to prevent the deterioration of, the quality of the natural environment"*. (Article I, based on Lundqvist 1996, 16).

The idea behind evaluation is simple, or "embarrassingly simple" as stated by Evert Vedung (2004, 2). Still, policy evaluation is a fairly recent phenomenon. Policy evaluation is mostly considered to have started in the 1960s (e.g. Furubo et al. 2002, Vedung 2004). It did not, however, take off at the same time in all policy areas or in all countries. The evaluation boom actually came to the environmental policy field only in the late 1990s (Knaap and Kim 1998a), which was also the time when it arrived more generally in Finland (e.g. Temmes 2000, Vedung 2001, Ahonen et al. 2002).

This thesis is thus about examining evaluation approaches, methods and criteria in an area where the evaluation demands are high, while concepts and practices lack any degree of standardization. The situation when the work on this thesis started is aptly described by the statement: "*Whereas the state of the art in program evaluation is in flux, the art of environmental program evaluation has no state at all. It has only artists."* (Knaap and Kim 1998b, 349) The underlying idea of the thesis is to combine knowledge accumulated in the general evaluation field with knowledge on environmental issues and policies and to test the derived concepts in practice. It is of course possible to conduct evaluations in total ignorance of the evaluation literature or any general evaluation approaches. If one is familiar with research methodology in general one might even do it well. It is also clear that each policy field, and each policy case has its own characteristics that must be taken into account when evaluations are conducted. Yet there are many issues where evaluators in one area can benefit considerably from the experience of evaluators in other areas and thus much can be gained in the field of environmental policy evaluation from the evaluation theory literature.

Next environmental policy will briefly be discussed (Section 2). Before then moving on to the concepts of environmental policy evaluation (Section 4) the evaluation demands will be examined (Section 3). The practical experiences of using the concept, in particular in the context of evaluating policies addressing emissions from pulp and paper production, will then be discussed (Section 5). The scope will then be enlarged by examining the evaluation of policy integration, that is, environmental policy goals implemented through policies with primarily other objectives (Section 6). Since a key aspect of evaluations is not just to carry them out but actually that they should be used, the use of environmental policy evaluations, conceptually as well as in practice, will be discussed (section 7). Before the concluding remarks (Section 9) the proposed concepts and their practical use will be linked to the evaluation specific debate between constructivists and realists (Section 8).

2. ENVIRONMENTAL POLICIES

2.1. WHICH POLICIES ARE ENVIRONMENTAL POLICIES?

There are many ways to separate environmental policy from other policies. Lundqvist (1996, 16) differentiated between approaches based on function, institution and purpose. An approach based on function views policies that affect the environment as environmental policy, whereas an institutional approach labels policies undertaken by a certain group of institutions – an environmental ministry, certain agencies, etc. – as environmental policy. Lundqvist advocated a purpose approach. He thus views environmental policy as *"courses of action which are intended to affect society – in terms of values and beliefs, action and organization – in such a way as to improve, or to prevent the deterioration of, the quality of the natural environment."* (Lundqvist 1996, 16)

The purpose-based separation of environmental policies from other policies, implies that not only all policies with environmental objectives as their main focus, but also policies with environmental aims incorporated among their key purposes are considered to be environmental policies. As stated in the introduction, as well as in Article I, a purpose approach is generally a reasonable one when there is a need to grasp what is meant by environmental policy. This approach will thus be used in this thesis unless explicitly stated otherwise. As will be discussed in Section 6, there are evaluation issues that require a different separation of environmental policy. A main objective of Agenda 21 adopted in Rio and incorporated in several European Union treaties is that sustainable development and environmental policy goals should be integrated into other policies. When evaluating policy integration, the purpose approach is not helpful, because any policy incorporating environmental objectives would then also become environmental policy. Evaluating policy integration based on a purpose approach, would actually mean that there would be no integration of environmental policy objectives into other policies, since according to the approach only policies that would not have integrated the objectives would be considered as other policies.

The separation of environmental policies from other policies is also linked to the issue of side-effects and unanticipated side-effects in particular. If what is considered environmental policy is based on function, that means that all policies that have affected the environment are taken to be environmental policies, then only environmental policies can have unanticipated environmental side-effects. This discussion may seem a semantic quibble. If ex post evaluations

are viewed as part of the feedback mechanism from the implementation of existing policies to the formation of new ones (Vedung 1997, 15-23), or if one considers the links between ex post and ex ante evaluations, it is, however more than semantics. A purpose-based separation of environmental policies from other policies, combined with a strong emphasis on unanticipated side-effects of any policy leads to different types of evaluations than a functional separation, not to speak of purpose or institutional separations without any recognition of unanticipated effects. The importance of unanticipated effects will be discussed in more detail in Section 4.2.

2.2. THE HISTORICAL BACKGROUND

In many countries examples of nature protection policies as well as action addressing particular environmental issues can be found since as early as the beginning of the last century. Still, most scholars agree that modern environmental policies started in the 1960s (e.g. Haila 1998). The book *Silent Spring* published in 1962 by Rachel Carson and translated into many languages is often mentioned as an important trigger (e.g. Haila 1998, 69).

In the late 1960s and early 1970s many industrialised countries adopted important environmental legislation and new organisations were formed that later became central for both policy development and implementation. For example, Japan passed its environmental protection law in 1967 (Andersen and Liefferink 1997, 4), the U.S. National Environmental Protection Act was enacted in 1970 (Davies and Mazurek 1998, 12) and environmental protection agencies were established in Sweden in 1967 (Kronsell 197, 40) and in the U.S.A. in 1970 (Weale 1992, 11).

In Finland several areas of environmental policy can be traced back much further than the 1960s. Nature conservation actually started – although on a small scale – back in the early nineteenth century, when Finland was still a part of Russia (Joas 1997, 121). The history of water pollution control can be traced back to the early 20th century, when the treatment of waste water became an issue in several cities (Laakkonen et al. 1999).

Even though the history of Finnish environmental policies has roots that go farther back in time it was from the 1960s that policies reminiscent of modern environmental policies started to emerge. Although water pollution control had been possible earlier, the Water Protection Act of 1962 was the first real piece of pollution legislation (Joas 1997, Article II). The scope

of Finnish pollution legislation has gradually expanded: in 1982 the Air Pollution Control Act came into force, in 1989 the Chemicals Act superseded the Poisons Act of 1969, in 1993 a New Waste Act became law, to mention only some of the important permit based legislation (Hildén et al. 2002).

In 1970, the first central environmental agency in Finland, the National Board of Waters, was founded and a separate Ministry of the Environment was established in 1983. An act passed in 1986 made environmental committees mandatory in municipalities with more than 3000 inhabitants. As part of the Finnish reform of central government the National Board of Waters and the Environment – the successor of the more focused National Board of Waters – was abolished in 1995. The administrative duties were largely transferred to the regional level, where thirteen Regional Environmental Centres were formed. At the same time the Finnish Environment Institute was created as a research and development organisation. (Hildén et al. 2002, 36)

In the early days of environmental policy, it was, in Finland as well as in other countries, mainly conducted through regulation, or "command-and-control" as it is often labelled (Davies and Mazurek 1998). Later, especially in the 1990s, the scope was broadened to include a variety of economic and information based instruments. Finland introduced the world's first CO_2 tax in 1990, legislation on environmental impact assessment was adopted in 1994, the EU's Environmental Management and Audit Scheme (EMAS) was enacted in 1994 (Hildén et al. 2002, 46) and the first energy conservation agreements were made in 1992 (Sairinen 2000, 228).

At the same time as the form of Finnish environmental polices has expanded from mere regulation to a wider use of different kinds of economic and information based instruments, the target groups and content also expanded. While the early policies focused mainly on point source pollution, such as water discharges from pulp and paper mills, to be discussed in detail in Section 5, emissions from non-point sources, such as agriculture and traffic are nowadays another important object of control. The target groups have also expanded from production alone to encompass consumption and use as well (e.g. Commission for Sustainable Consumption and Production 2005). Although emissions are still an important part of the policy content they are complemented with emphasis on the use of natural resources and energy, chemical risks and biodiversity protection, to mention only a few of the most important new focuses.

In parallel and linked to the development of national environmental policies, international

policy formation has emerged. Already in the beginning of the 1970s OECD adopted the Polluter Pays Principle and the Declaration on the Human Environment was signed at the United Nations conference in Stockholm in 1972. The Rio Declaration on Environment and Development as well as some new conventions, such as the UN Framework Convention on Climate Change, followed in 1992. (Mickwitz 1998) Although the European Union had no environmental policy when the European Economic Community was founded in 1957, the environment is, since the Single European Act that entered into force in 1987, included in the treaties (Hildebrand 2002). The EU is on its sixth environmental framework programme and the number of environmental regulations, directives and decisions adopted has increased. From 1993 to 1995 on average 48 environmental regulations, directives and decisions were adopted or amended annually (Zito 2002, 160). Since Finland agreed to join the European Economic Area in 1992 and subsequently joined the European Union in 1995 the content, structure and style of Finnish environmental policies are greatly influenced by EU policies. Policies agreed upon in regional organisations, such as the Baltic Marine Environment Protection Commission (HELCOM) or the United Nations Economic Commission for Europe have also had important links to the policy development of their member countries.

2.3. SHARED CHARACTERISTICS OF ENVIRONMENTAL PROBLEMS

Although every environmental problem has it own case- and context-specific characteristic many have argued that they often have some features in common (e.g. Weale 1992, Article I). It has also been argued that these shared features make these problems particularly difficult to solve (e.g. Lafferty and Meadowcroft 1996, Weale 1992). A key argument of this thesis is that these features should be considered when evaluations of environmental policies are planned and carried out (Articles I, III and IV). The context-specific characteristics of any environmental policy evaluation setting, of course, also have to be taken into account and not all of the common characteristics to be discussed here may be of relevance in all cases. The shared features could thus be seen as a priori assumptions, which are useful to consider but might be dropped after consideration in individual cases.

Lafferty and Meadowcroft (1996) argue that the following features characterise environmental problems and make them hard to handle: the knowledge deficit, which includes the element that the problems often are complex, technically difficult and uncertain;

complex geographical patterns of impact and causation, the problems often link regions that are geographically but also socially remote; redistribution of loss and gains, often any action on an environmental problem will result in large redistributions of costs and benefits; and time-scale effects, the problems are often due to activities that have been going on for very long and their full effects will often become transparent only in the distant future and the results of any action against them will also often not be seen until a long time has passed.

Weale (1992, 9) argues that pollution control, which is a subset of environmental policy, has the following characteristics: it is a matter of providing public goods; it involves complex technical issues; the timescale is often very long; and it cuts across sectors if solutions are sought at sources rather than by dealing with effects.

Others have added to the lists of Lafferty and Meadowcroft and Weale or formulated the features differently, but there is a fairly broad consensus on most of the features. Based on these assessments of typical characteristics of environmental problems, Table 1 summarises key features that are also essential for the evaluation concepts one could use when evaluating the policies formed to address these problems.

Table 1. Summary of the key characteristics of environmental problems under consideration (Article I)

Features of the problems	Features related to the knowledge
• They are complex	• They have been formulated as problems largely by scientists
• They have long time frames	• They involve huge uncertainties
• They concern geographically remote areas	• They involve stakeholders with different belief systems and conflicting goals
• Their consequences and causes are unequally distributed	

In order to exemplify the role of these features they will be discussed briefly for two environmental problems. Through the discussion about the two cases a glimpse of the case to case variation is also provided. One of the problems, climate change, has been the most publicly discussed environmental policy issue lately. The other one issue, industrial water pollution, is not an environmental problem itself, but one of the causes to several environmental problems of which eutrophication is one. Industrial water pollution has been the empirical case of two of the Articles (II and IV).

The physical processes behind climate change are complex and uncertain and when the economic, social and political aspects are included the complexity and the uncertainties grow even further. For the main gases that contribute to climate change it is estimated to take over a

decade before an addition of methane is removed from the atmosphere, while carbon dioxide (CO_2) may persist up to 200 years and the lifetime in the atmosphere of perfluoromethane has been assessed to even more than 50,000 years (IPCC 2001, 38). While the time taken for temperatures to rise as a result of higher concentrations of CO_2 may be more than a century the sea level responds even more slowly, due to the large heat capacity of the oceans. Thus sea-level rise is expected to continue for centuries even if the climate were to stabilize (IPCC 2001). The long-term frame of many climate change related processes implies that the final outcomes of actions to counter climate change will not be detectable for a long time.

Any action against climate change will be largely local but the effects of climate change will be global, although different regions may be affected in very different ways. Developing countries are often expected to be more adversely affected by climate change than developed countries, although changes, for example, in the Gulf Stream could affect the expected regional distribution of the effects. Most developing countries are in tropical regions, are more dependent on agriculture and natural resources and have fewer resources to adapt to climate change (Banuri et al. 1996, 97). Not only the consequences of climate change are expected to be unequally distributed; so are the causes. In 2000, the per capita CO_2 emissions in the US were 20.6 metric tons, the Finnish per capita emissions were 12.0 metric tons, while those in India were only 1.1 and in Ethiopia a mere 0.1 metric tons per person (United Nations 2005). This is largely because the energy intensity of the economies varies so much; for example, the average American uses 16 times as much energy as the average inhabitant of India (Economist 2003, 94).

Nobody who has followed the public debate is unaware of the knowledge conflicts related to climate change. Although it seems reasonable to argue that there is nowadays a broad consensus among scientists that climate change caused by human activities is taking place, even this is not undisputed. While the enormous research efforts related to climate change have reduced some uncertainties, new unknowns have emerged (Skodvin 2000, 177). The expected effects of different policy instruments not only on emissions, but even more on the innovation and diffusion of technologies is a clear example of an area in which different actors have different belief systems.

Although the processes caused by water discharges are not as global as climate change, there are several complexities and many aspects are uncertain. There has long been a debate among scientists over the limiting role of phosphorus versus nitrogen and recently a Swedish report argued that the Baltic Sea might be in a new equilibrium, implying that there is no return to a state resembling that of pre-industrial times (SOU 2005).

Any action against water discharges will mainly be local, although parts of a water treatment technology may be imported from the other side of the world and thus provide work and maybe also cause environmental degradation where the technology is produced. Many pollutants from activities inland in Finland will eventually have effects on the whole Baltic Sea, although different regions may be affected in very different ways. There are clearly different perceptions of the role of runoffs from Finnish agriculture versus, for example, waste water from St. Petersburg. The present eutrophication of the Baltic Sea has been affected by the discharges of phosphorus and nitrogen over many decades. Due to accumulation and internal loads the effects of policies now to reduce the discharges, even if implemented rapidly, will not be seen for decades. (Article I)

This brief discussion of climate change and eutrophication caused by industrial waste water discharges showed many similarities, but also differences. Climate change is a more complex problem, the time frame of the environmental processes is longer and in being a global phenomenon it concerns all peoples, regions and countries. The waste water discharges from Finnish industry have local effects on the water courses into which the waste waters are discharged, but often also affect the Baltic Sea. While there are equity issues related both to the different countries as well as different sectors within countries the differences in means are not as large as in the case of climate change, where the poorest and the richest people on earth are affected. Similarly the uncertainties and conflicts in belief systems are also larger in the case of climate change. While there has not been any major scientific conflict lately about whether industrial waste water discharges contribute to eutrophication, the debate has rather been about the relative cost-effectiveness of reductions in different sectors and countries.

3. PRESSURES TO EVALUATE ENVIRONMENTAL POLICIES

3.1. FOUNDATIONS OF THE DEMANDS TO EVALUATE ENVIRONMENTAL POLICIES

Why evaluation – including environmental policy evaluation – should be undertaken is mainly justified by two separate functions, namely learning and development on the one hand and accountability on the other.[1] It is through these functions that the demands for evaluations are justified. Accountability is about the liability of those, organisations as well as individuals, entrusted with public tasks and resources. Accountability has been extended from ensuring that resources are not misused, which is the task of auditing, to making sure that the resources are used wisely and that the stated goals are achieved.

Learning and development through evaluations is a very important and broad issue. The opportunities to learn from evaluations are numerous (Article I). Those participating in the political process through which policies are formed can learn, likewise those implementing the policies. Learning can be based on the evaluation findings, but more often the opportunity to learn from the evaluation process is emphasized. Evaluations may provide opportunities to learn about the questions to ask, the goals to set and how to frame the issues as well as the instrumental learning about how to design or implement the policy.

Learning and development can take place at many levels; individuals can learn but so can organizations. Single and double loop learning (e.g. Argyris 1999, Leeuw et al. 2000) are useful distinctions; in single loop learning an organisation develops its practices so as to perform its task more successfully while double loop learning means that the tasks and the fundamental conditions for accomplishing them are also re-examined. Argyris (1999, 68) uses the example of a thermostat; in the case of single loop learning, it would react to the information "too cold" by turning on more heat and to "too hot" by turning off the heat. A thermostat that would ask itself why it is set at a particular temperature and re-examine the other conditions affecting how it had been programmed would be engaged in double loop learning.

1 For what evaluations are actually used once carried out or after they have been conducted is not necessarily the same as the functions by which evaluations are justified when requested. The issue of use and different types of use is discussed in Section 7.

The role of trust and legitimacy in government and organizations implementing public policies is emphasized particularly in the recent debates concerning social capital. At the same time, however, there is a trend towards diminishing trust in and decreasing legitimacy of the state. Although in Finland these problems have not been as severe as in many other countries (Bouckaert et al. 2000, 14) it is still important to ensure trust and legitimacy and policy evaluation is frequently justified through the task to ensure that elected bodies and implementing organisations are accountable. Following Peters (2001, 300), *"In its simplest form accountability means the requirement of a public organization (or perhaps an individual) to render an account to some other organization and to explain its actions."*

Analytically one can view public policy through a series of principal agent relationships (Vedung 1997, 2004). Principal agent relationships are based on delegation; instead of conducting everything himself a principal transfers one or several tasks to one or several agents. Principal agent relationships represent interactions characterised by a lot of imperfections, for example in information, control and verification. In a representative democracy the chain starts from the citizens being principals who elect representatives, e.g. parliamentarians, i.e. their agents. These are then the principals of the government, but the chains continue through national, regional and local agencies implementing policies all the way to the "street level bureaucrats" who deliver the service. It can be argued that the longer and the less transparent the chain becomes the greater is the need to guarantee accountability through evaluation. The frequent requests to evaluate international and EU policies correspond to this view.

3.2. STATED DEMANDS TO EVALUATE ENVIRONMENTAL POLICIES

The European Union is nowadays requiring evaluations of environmental policies. This is however, a fairly recent phenomenon. In the 5th environmental framework programme there were still no articulated demands for policy evaluations. The assessment of this framework programme stressed the need for policy evaluations. For example, it stated that *"The effectiveness of environment policy should be regularly assessed and remedial measures taken if required."* (European Commission 2000, 19) In its conclusions based on the assessment the Environment Council stated that the lack of *"a systematic ex-post evaluation process, appropriate monitoring mechanisms and indicators, does not allow a thorough assessment of the effectiveness, in terms of reducing environmental impacts and risks, of different Community environmental policy*

measures" (European Environment Agency 2001, 9). Several non-governmental organisations such as the Union of Industrial and Employers' Confederations of Europe and Birdlife, had also stressed the need for policy evaluations in their comments to the assessment (European Commission 1999).

Based on the discussions about the 5th framework programme, the task of evaluating policies was clearly stated in the 6th Environmental Action Programme for the European Union (1600/2002/EC), which was adopted in June 2002. Paragraph c in Article 10 states:

"[The objectives shall be pursued by] *improvement of the process of policy making through:*

- *ex ante evaluations of the possible impacts, in particular the environmental impacts, of new policies including the alternative of no action and the proposal for legislation and publication of the results;*
- *ex post evaluation of the effectiveness of existing measures in meeting their environmental objectives."* (European Parliament and the Council of the European Union 2002)

The present attitudes towards the evaluation of environmental policy are well captured in the foreword by Lars-Erik Liljelund (2004) to the European Environment Agency's strategy 2004–2008. He states: *"I would like to highlight two strategic dimensions of the further development of the EEA … Secondly, the increased emphasis that will be placed on evaluations of policy effectiveness. Environmental policy is no longer a free ride. In order to be able to convince politicians and the public alike that environmental policies are necessary and good for society as a whole, we must be able to demonstrate that they are delivering real results in an effective way."*

The demands to evaluate environmental policies are not only expressed at the EU level, but also in many countries. For example, in Sweden one of the key tasks of the Environmental Protection Agency is evaluation. The brochure presenting the Swedish EPA states: *"Evaluations are an important basis for further development of environmental policy and this is one of the areas to which the Swedish EPA is giving highest priority."* (Swedish EPA 2000, 5) In Denmark a new institute "the Environmental Assessment Institute" was established in 2002. According to the news reports the institute was established in order *"to enable Denmark to get 'more environment for the money"* (ENDS, January 15 2002). The objective of the EAI is, on the basis of *"research at a high, international level, to contribute to reaching environmental objectives in the most economically effective way"* (Environmental Assessment Institute 2005). The institute, as well as its first reports, was very controversial. That Bjørn Lomborg was the first director of the

institute, until August 1st 2004, increased the controversy. But also outside of the EU, a similar development can be seen. For example, at the US Environmental Protection Agency (EPA) an Evaluation Support Division, was established in the Agency's policy office, in 2000 (Swett 2004). This was a response to external as well as internal pressures for a renewed interest in policy evaluation in EPA

In Finland, too, the interest in evaluating environmental policies has gradually grown. According to a study conducted in 1998 the interest in evaluating the effects of legislation was lowest in the Ministry of Justice and in the Ministry of the Environment (Harrinvirta et al. 1998). Since 1998, however, the situation has changed. In 2001 an evaluation plan for the environmental administration up to 2006, was adopted (Ministry of the Environment 2002, 32). In the strategy published in 2002 it is stated that *"In our opinion it is important that the effects of decisions and actions are evaluated both in advance and afterwards and that actions are cost-effective."* (Ministry of the Environment 2002, 8) Evaluations of environmental legislation is listed as an indicator for the objective "improving the effectiveness of policy instruments", which is defined as one of seven key tasks for the Finnish environmental administration up to 2009 (Ministry of the Environment 2004).

Since the end of the 1990s the Ministry of the Environment has commissioned several policy evaluations and the work on this thesis was also started as part of a response to an evaluation request by the Ministry of the Environment, (also Articles I, II, III and IV). Furthermore, not all evaluations of Finnish environmental policies originate from the Ministry of the Environment. For example, the State Auditors have undertaken a couple of evaluations of environmental policy (e.g. State Audit Office 2001, 2002).

3.3. FRAMING THE ENVIRONMENTAL POLICY EVALUATION DEMAND

The recent requests to evaluate environmental policies have to be seen in context. The relevant context can be divided into two interlinked parts; the general public policy context and the environmental policy specific context. While the demands to evaluate environmental policies have occurred during a time characterised by a general scepticism in public policy[2], new

2 A good illustration of this attitude is the statement by President Clinton "The era of big government is over" (Gore 1996, 1).

public management and liberalisation, this does not mean that requests for evaluations have not developed in another context. Quite the reverse, when evaluation first emerged in several sectors in the so-called pioneer evaluation countries, e.g. USA and Sweden, it was in the 1960s, an era when there was a deep trust in planning, big societal programmes and reforms. Considering the evaluation demands expressed in their policy context, such as the "new public management", implies that their form, the expectations created, but also the response and finally the use of the evaluations may be very different in another context.

Public sector reforms have swept over all OECD countries especially since 1980 (Pollitt and Bouckaert 2000, 2004). Many of these reforms have been conducted under the umbrella of "new public management". According to Rhodes (2000, 56) new public management originally consisted of two main ideas; corporate management and marketization, where corporate management refers to the effort to import management practices from the private sector to the public sector. These practices included, for instance, performance measurement, management by result and customer focus. Marketization on the other hand refers to redefining the tasks as well as the way they are provided, through contracting-out, privatisation, creation of quasi-markets, etc.

In Finland, too the public sector has been reformed based on ideas contained in new public management (e.g. Temmes 1998). The Finnish reforms are characterised by *"a history of substantial but nondoctrinaire reforms, which have been adopted calmly and continuously"* (Pollitt and Bouckaert 2004, 49). As a whole these reforms have resulted in the public sector being remarkably different in the late 1990s compared to what it was a decade earlier (Temmes 1998, 449). Some of the key features of the Finnish reforms are: a sharp decrease in the number of civil servants, through the creation of State Enterprises and in some cases later State-Owned Companies; the introduction of a framework budgeting system and results-oriented budgeting; restructuring of central agencies and shifting responsibilities to the regional and especially the municipality level; and reform of human resource management including the salary systems (Pollitt and Bouckaert 2004, 241-2).

The Finnish public sector reforms have produced several managerial and administrative challenges that accentuate the demand for evaluations. The administrative culture has developed from a dominantly administrative-legalistic one into one where public interests are more emphasised (Pollitt and Bouckaert 2004, 54, Temmes 1998). The functional specialisation of administrative sectors and the greater independence of regional and local administrations have increased their need to consider and report economic as well as other

impacts of their activities. Result-oriented budgeting calls attention to demonstrate effects of policies. Directing the whole governmental apparatus and achieving co-ordination, without blurring the independence and customer responsibilities of individual agencies and public enterprises has become a huge challenge (Temmes 1998, 452) Finally, the evaluations undertaken of the public sector reforms themselves (Pollitt and Bouckaert 2004, 243) and of new or remodelled institutions (Ahonen et. al 2002) have been important for the general development of evaluation capacity in Finland.

Many of the general reforms of the Finnish public sector have also affected the environmental administration (Cederlöf 2001, 155-6). As already discussed (Section 2.2), the central environmental administration was reformed in 1995 and many duties have been transferred to regional agencies and municipalities. The new budgeting approaches naturally also concern the environmental administration, both in relation to the state budget and the Ministry of Finance as well as internally. Achieving co-ordination, within the environmental administration but also between all policies affecting the environment, including policies formed and implemented by other sectors, has become a major challenge (Section 6). As of 2004 new result based salary systems have gradually been introduced into the environmental administration. The Ministry of the Environment has the overall responsibility for environmental policy, but its means to influence and co-ordinate the activities of the regional agencies and municipalities are substantially different than earlier.

The evaluation demands can also be examined by comparing the development in different countries. Furubo and Sandahl (2002) have used a categorisation based on internal and external pressures for evaluations to analyse how evaluations has developed in different countries. The main source of external pressures to develop evaluations is the EU, but also other international organisations such as OECD and the World Bank, whereas the internal pressures relate to a wide range of domestic needs from requirements of more transparency to public savings. The development of an evaluation culture in Finland was categorised by Furubo and Sandahl (2002) to be due to both strong external and strong internal pressures.

The demand for evaluations of environmental policy is thus partly due to new public management, lack of trust in the public sector and an – at least perceived – lack of public resources. However, there is, also a range of environmental policy specific factors. Since many environmental policies were adopted in the 1960s and 1970s they have been in force long enough to make learning from the experiences possible. In the European Union the number of environmental directives and regulations has increased (Section 2.2.) as have the

acts and decrees in many countries. There is thus more to evaluate and potentially more to be gained from evaluation. Many proposed policies, such as the EU Commission's proposed new chemical legislation REACH (Registration, Evaluation and Authorisation of CHemicals) or policies addressing climate change, are supposed to have very widespread effects and have thus induced many ex ante evaluations. The Aarhus Convention on access to information and participation in environmental matters also provides a basis to use evaluations to make environmental policies more transparent. Finally the request on environmental impacts assessment of projects as well as programmes and strategies can be, and has been, applied also to demand impact assessments of programmes and strategies of environmental policies.

4. CONCEPTS OF ENVIRONMENTAL POLICY EVALUATIONS

4.1. POSITIONING ENVIRONMENTAL POLICY EVALUATION

As already stated in the introduction, policy evaluation is in this thesis considered to be: *"careful assessment of the merit, worth and value of administration, output and outcome of environmental policies, which is intended to play a role in future, practical action situations."* This is the definition by Vedung (1997, 3) with the exception that it is not limited to retrospective assessment as in Vedung's definition. Most evaluation theorists as well as practitioners nowadays recognise ex ante, intermediate as well as ex post evaluation (Chelimsky 1995, Nagarajan and Vanheukelen 1997, 72). Pre-evaluations are particularly important in the field of environmental policy, since one type of ex ante assessment – environmental impact assessment – is an important aspect of environmental policy in many countries. Requirements for ex ante environmental impact assessment have been extended from projects to policies and programmes as well. Although the definition used is extended from just retrospective evaluations to pre-assessment as well and most of the concepts proposed could be used for both types of evaluations, the discussion on the practice of using these concepts will be limited to ex post evaluations. In that sense Vedung's (1997, 3) original definition is actually applied.

The definition of policy evaluation starts out with "careful assessment". Some authors, e.g. Rossi et al. (1999, 4) view evaluation as a particular use of social research, i.e. evaluation is a form of research. While evaluation can be research, the phrase "careful assessment", is used to make it broader. This distinction can be demonstrated by an example; peer review. Peer review is often used to evaluate the quality of academic work, but it is also frequently used to evaluate institutions. The Finnish Environment Institute was evaluated through peer review in 1998 (Ministry of the Environment 1998). Peer review is based on the idea that the work of a professional, or an organisation, in a field can be assessed by somebody else in that same field. The emphasis is not on how the evaluation is done but by whom. While a peer review is hopefully careful it can seldom be called research.

Since the seminal work by Pressman and Wildavsky (1984) in the 1960s and 70s, implementation studies became an important field of policy research. It is obvious that implementation studies are closely related to policy evaluation. Implementation can be seen as "what happens" between policy formulation or expectations and policy results (Hill and Hupe

2002, 2). In the wording of the evaluation definition above, implementation would be the administration and the output of environmental policies. There is thus a difference between implementation and evaluation. *"Implementation providing the experience that evaluation interrogates"* (Wildavsky 1984, xv) but there is interaction in the other direction as well, since in a dynamic perspective *"implementation is … about learning from evaluation"* (Browne and Wildavsky 1983, 204). If evaluation and implementation are interacting so closely, does it mean that implementation studies and evaluation studies are actually the same thing? Hill and Hupe (2002, 11-12, 141) start from the notion that the perspectives of implementation studies and evaluations are different, but they conclude: *"In a sense implementation studies are not so much different from evaluation studies as a sub-set of them"*. The view of implementation studies as a sub-group of evaluations, focusing in particular on the administration and output of the policies, is consistent with the view of this thesis.

The use of the phrase "administration, output and outcome of environmental policies" in the definition of environmental policy evaluation incorporates several different types of evaluations. First an evaluation could only focus on the administration or the merit, worth and value of the administration in generating outputs, e.g. environmental permits. More frequently, however, the declared evaluation demands are about the policies in relation to the outcomes. Policy evaluations may be focused differently, or have different "evaluands" (Scriven 1991, 139) in the evaluation jargon. Examples could be programmes, for example the Finnish climate change programme; specific legislation, such as the Water Act; a specific part of an act or a decree, e.g. permits or even a particular value in a decree; a policy strategy potentially implemented through several programmes and laws, e.g. the Finnish programme on sustainable development; or an institution, such as the Southeast Finland Regional Environment Centre. Although the practice of the discussed concepts is from the evaluation of policy instruments they could be applied to the other evaluands as well (Section 4.6).

4.2. SIDE-EFFECT EVALUATION

The European Union, like many countries, is focusing its demands for environmental policy evaluation largely on evaluating the effectiveness or the cost-effectiveness of policies (Section 3.2). The term effectiveness generally means, *"to what extent have the intervention's impacts contributed to achieving its specific and general objectives?"* (Nagarajan and Vanheukelen 1997,

71, EEA 2001, 9). While evaluating the effectiveness of policies and programmes, including those in environmental policy, is necessary, it is not sufficient (Article III). It is necessary, because objectives are key elements used in the policy formation process in order to get policies adopted and thus they should be followed up. This is the case both for objectives determined by politically elected bodies or by public agencies exercising the mandate they have been given. This is simply a requirement of a transparent democratic process. To only assess the effectiveness, however, is not enough; a broader perspective based on several criteria and also considering side-effects is required. Only then can meaningful inputs to the deliberation on the *merit*, *worth* and *value* of a policy be provided through evaluations. Side-effects will be discussed in this sub-section and criteria in the next.

Effectiveness evaluation can partly be seen as a reappearance of the oldest evaluation model; the "goal-achievement model" under a new name (Article I, III). The rationale behind this evaluation model is simple; it is based on the question "are the results in line with the goals?" followed by the question "are the results due to the evaluand?" (e.g. Scriven 1991, 178). The most obvious problems with the goal-achievement model are that it disregards side-effects and unanticipated effects; it does not consider costs; and the relevance of the goals is not questioned.

It is well established that public policy seldom turns out exactly as intended. This is partly due to implementation; what is delivered is not what was expected. It can partly be caused by policies being based on incorrect assumptions (Hoogerwerf 1990). But it is also because policies, especially important and interesting ones, tend to be used in complex and changing contexts, where there are many other actors as well as external factors and the interactions between these are uncertain. The effects of policies, even when extensively planned, are therefore often unanticipated. *"Social life is not only a trial of strength between opposing groups: it is action within a more or less resilient or brittle framework of institutions and traditions, and it creates - apart from any conscious counter-action - many unforeseen reactions in this framework, some of them perhaps even unforeseeable."* (Popper 2003, 105) Popper (2003, 105) therefore even argues that "*the main task of the social sciences … is the task of analysing the unintended social repercussions of intentional human actions."*

The "side-effects evaluation" model (Vedung 1997) is a response to the criticism about the limitations of goal-achievement or effectiveness evaluations, and as argued in Article I, this is an approach often especially suitable for the evaluation of environmental policies. In this evaluation model the effects of the evaluated policies are first divided into anticipated and unanticipated

effects. Then they are categorized based on whether the effects occur inside or outside the target area and finally there is a qualitative categorization of the effects[3] (Figure 1).

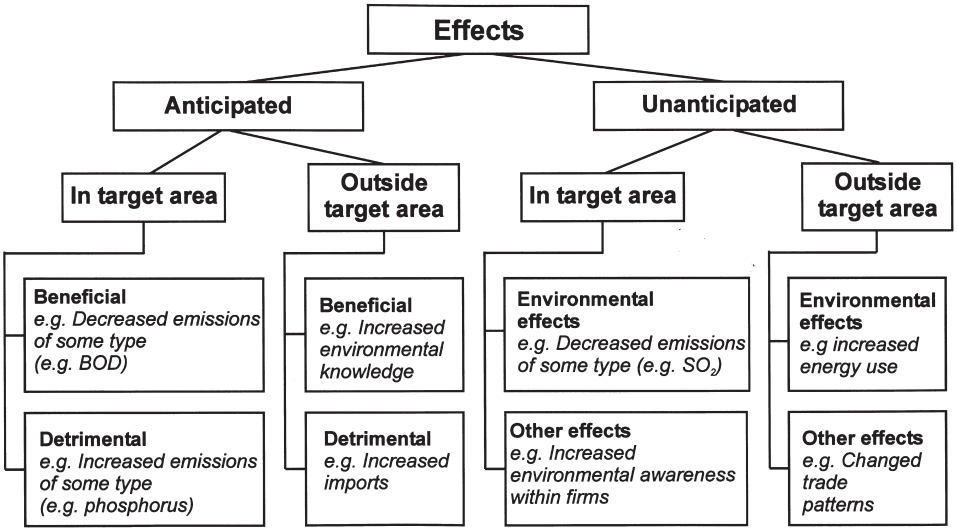

Figure 1. Different types of effects: classification and examples (Article I)

Since many environmental policy problems are complex, negative environmental effects, either inside or outside the target area, can often be expected and sometimes anticipated. Due to the huge uncertainties frequently involved in environmental problems, environmental policies as well as other policies tend also to have important unanticipated environmental effects. Environmental policies that have reduced one problem have often achieved this by: shifting the problem regionally through for example "higher chimneys"; increasing other problems, e.g. pollution of other substances; or by postponing the problem (e.g. Jänicke and Weidner 1995). Since environmental policies not only affect the environment, but many other processes and aspects of society as well, and because these relationships, e.g. social exclusion, are also complex and uncertain, unanticipated environmental effects are not the only potential unanticipated effects.

3 An important issue is of course to what extent a simple dichotomy, such as "beneficial" versus "detrimental" is useful. What is beneficial is often heavily dependent on whose perspective is taken. The anticipated effects can often be classified in this way based on the preparation material for the policy. Classifying an effect as beneficial does not imply that it cannot be problematised. Actually one reason for using several criteria (next Section) is to assist a systematic but also critical treatment of all kinds of effects. It is often not important to divide all effects into "beneficial" and "detrimental", especially not the unanticipated effects. Instead it is often sufficient merely to divide unanticipated effects into environmental effects and other effects (Figure 1). (Article I)

If an ex post evaluation is conducted long after a policy was introduced some effects that were not anticipated when the policy was formulated may be known when the evaluation is started. The unanticipated effects are often still only partially known before the evaluation is actually carried out. One important task, sometimes even the most important task, of an evaluation is therefore to make the picture of the unanticipated effects of the environmental policy evaluated more complete. (Article I)

4.3. MULTI-CRITERIA EVALUATION

Evaluation is by nature normative and thus some criteria on which to base the normative judgements must be utilised. The choice of criteria can be based on two general approaches: a descriptive and a prescriptive one (Shadish et. al. 1995, 47-8). The evaluators select the criteria in the prescriptive approach, whereas the criteria are based on goals or needs stated by others, e.g. by legislators or stakeholders in descriptive approaches. Another feature related to the choice of criteria is when the criteria to be utilised in an evaluation are decided. Criteria could be decided before the evaluation is started as part of the evaluation plan. But often new features of the evaluated policy are revealed during the evaluation process and these may necessitate additional criteria. Finally, those commissioning an evaluation and other users might want to utilise additional criteria after the evaluation is completed. The option for others to use additional criteria requires that the evaluation process and documentation, e.g. materials and methods, are transparent.

The discussion about multiple criteria (Table 2) in Articles I, III, IV, had two objectives; first to demonstrate that by utilising more than one single criterion, be that effectiveness or any other, a more comprehensive debate about the policy is facilitated and second to suggest some criteria that could form a starting point in context specific evaluations. The criteria discussed are far from the only ones possible and they would have to be specified for any particular situation. They can thus assist an evaluation both when it is based on a prescriptive or a descriptive criteria specification and selection. In Article I, three groups of criteria are discussed: general criteria, economic criteria and criteria linked to the functioning of democracy. In Article IV, the criteria are grouped into three categories: criteria generally requested, other general criteria and criteria related to democracy.

Impacts and effectiveness have already been discussed at some length in the previous subsection and since the other criteria are discussed at length in Articles I an IV only some criteria will be briefly discussed here.

Table 2. Criteria that can be used in evaluations of environmental policies and examples of their interpretation. (Based on Articles I and IV)

Criterion	Related questions	Group
Impact	Is it possible to identify impacts that are clearly due to the policy and its implementation? All impacts may be considered in the light of this criterion, irrespective of their occurrence within or outside the target area or whether they are anticipated or not.	A generally requested general criterion
Effectiveness	To what degree do the achieved outcomes correspond to the intended goals of the policy?	A generally requested general criterion
Relevance	Do the goals of the instruments cover key problems of environmental policy?	A seldom requested general criterion
Flexibility	Can the policy instrument cope with changing conditions?	A seldom requested general criterion
Predictability	Is it possible to predict the administration, outputs and outcomes of the policy instrument? Is it thus possible for those targeted, as well as others, to be prepared and take into account the policy and its implications?	A seldom requested general criterion
Persistence	Are the effects persistent in such a way that they have a lasting effect?	A seldom requested general criterion
Efficiency (cost-effectiveness)	Do the results justify the resources used? This is a cost-results criterion, in which benefits are not valued in monetary terms. Another possibility of how to consider costs is to use the cost-effectiveness criterion: Could the results have been achieved with fewer resources?	A general economic criterion
Acceptability	To what extent do individuals and organisations accept the environmental policy?	A criterion related to democracy
Transparency	To what extent are the outputs and outcomes of the environmental policies, as well as the processes used in the implementation, observable for outsiders?	A criterion related to democracy
Participatory rights	Who can participate in the processes through which the environmental policies are implemented?	A criterion related to democracy
Equity	How are the outcomes and costs of the environmental policy instrument distributed?	A criterion related to democracy

Because environmental problems are often characterized by the conflicting goals of different groups and by huge uncertainties even in the causal relationships, a comprehensive treatment of relevance is particularly important. Complexity and non-linear dynamic relationships characterise the context of many environmental policies, there are therefore often many external factors affecting outcomes and these often change over time. The flexibility of environmental policies is therefore often an essential feature. Flexibility may, however, be in conflict with other criteria, especially predictability but partly also persistence.

Assessing the costs, and sometimes also the benefits, of policies in monetary terms can be viewed as utilising a specific evaluation model (e.g. Vedung 1997, 86). One can, however, also argue, as in Article I, that although such approaches require special concepts and models they can be viewed as economic criteria and thus linked to a multi-criteria approach. Applying economic criteria has the feature that many, or even all aspects, are calculated in terms of the same unit – money – which makes it easy to develop clear decision rules (e.g. Braden and Kim 1998, 210). The procedure of giving all effects monetary values may make explicit assumptions and judgements that are often otherwise hidden.

Utilising economic criteria, however, is almost always problematic, both conceptually and practically. The conceptual problems are related to both aggregation and valuation. Although economists have developed several methods to create monetary values for effects that have no market value, they are all still associated with problems. The downside of aggregation is that it conceals the details, such as whose costs, when they arise and where. The practical problems associated with economic criteria are often related to data availability (e.g. Braden and Kim 1998). Recognising these problems, conceptual as well as practical ones, does not mean that economic criteria should not be used. Ignoring costs in a world preoccupied with limited resources would be absurd. The implication is rather that these problems have to be handled in explicit and transparent ways and that efficiency in a cost-effectiveness sense is in practice often an easier and more usable criterion than efficiency in a cost-benefit sense. (Article I)

Lafferty and Meadowcroft (1996) among others have argued that environmental problems constitute particular problems for democracy. Criteria such as acceptability, transparency, participatory rights and equity, which are important in any policy evaluation, are given a particular status in evaluating environmental policies against this background. While these criteria are important on their own they are also related to the other criteria since acceptable and transparent policies might also have greater impacts and be more effective.

In the case of environmental policy equity is very important, both with respect to outputs, outcomes and costs and with respect to the process by which policies are formed and implemented (Banuri et al. 1996). The discussions about equity too often focus only on outputs and outcomes, including costs. The equality of the process, covering aspects such as equal participation and equal access to information, should, however, not be overlooked. Especially since the capabilities of stakeholders often vary a lot, both internationally and nationally, equal possibilities often requires explicit recognition in the process design to come about. The long time scales linked to many environmental issues emphasise the distribution over time, i.e. intergenerational equity, of effects and costs, in addition to the present distributions (e.g. Portney and Weyant 1999).

A feature of multi-criteria is that aggregation into a simple statement of the policy such as "good" or "bad" is difficult. This can bee seen as a problem but also as a strength. If one supports the views expressed by Shadish, Cook and Leviton (1995, 46-52) and Vedung (1997, 248-249) that pluralism of values should be appreciated in democratic societies a lack of a simple aggregated "verdict" can be viewed as a good thing. The multiple views of the policy would facilitate a communicative political process rather than restrain it.

4.4. INTERVENTION THEORIES

Since Huey-Tsyh Chen (1990) published his book "Theory-driven Evaluations" in 1990 programme theory, or the more generic concept intervention theory, has become a central topic in the discussions on evaluation methods. The concept, however, is much older than the 1990s. The idea that evaluations should test the "chain of objectives" can be traced back to the 1960s (Rogers et al. 2000, 6). Frans Leeuw (2003, 6) even traces the idea back to the sociologist Karl Mannheim in the 1930s.

Inspired by Andries Hoogerwerf (1990), Evert Vedung (1997, 301) has defined programme theory or intervention theory, as: *"All empirical and normative suppositions that public interventions rest upon"*. An intervention theory can be seen as a model *"of the microsteps or linkages in the causal path from program* [or more generally intervention] *to ultimate outcome"* (Rogers et al. 2000, 10) on the basis of the detailed assumptions of how the intervention is intended to work. The role of intervention theories is to describe how a policy is supposed to be implemented and function. Intervention theories are not descriptions of how an intervention

actually works, but are rather to be used as tools in the evaluation process in order to assess the actual implementation and the effects the policy has had in practice. (Articles I, III and IV)

It is rather a rule than an exception that an intervention is based on several intervention theories, not just one, since different groups often have different expectations of an intervention (Weiss 2000). An evaluation using intervention theories should therefore reflect these different assumptions by constructing multiple intervention theories instead of just one. (Articles I, III and IV).

Figure 2. provides an example of what an intervention theory, or actually part of an intervention theory, for Finnish waste water permits may look like. This example of an intervention theory is a sketch, since usually a far more detailed intervention theory is required. While the example in Figure 2 is a rather linear intervention theory (it only has two feedback loops) it should be stressed that there is nothing in the concept of intervention theories that limits them to linear ones.[4] An alternative intervention theory to that in Figure 2 could start from the firm's current technology and be based on the assumption that the permit limits are based on technology already in use and thus no new investments would be induced. Neither would reductions of discharges follow from the permits; they would merely be an insurance against increasing discharges.

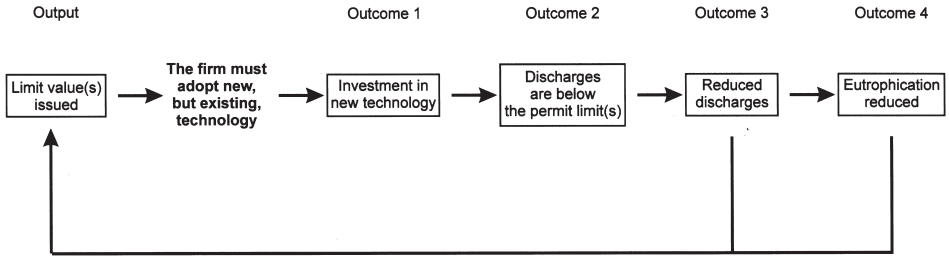

Figure 2. Part of an intervention theory for Finnish waste water permits (Article I)

4 There is also no contradiction in principle between a linear intervention theory and a complex and non-linear world. An intervention theory is not a description of the policy intervention in the real world, but of the assumptions of the intervention and these assumptions may be linear even if the practice is not.

Some essential notions that are useful when restructuring an intervention theory are listed in Table 3.

Table 3. Some key elements of intervention theories (Based on Articles I, III and V)

Notion	Explanation
Target groups	The actors, i.e. decision making entities, such as companies, other organisations and individuals whose actions the policy is intended to influence.
Outputs	The assumptions about what the administration produces and the target groups are faced with, e.g. a seminar, an environmental permit with detailed conditions or a subsidy paid. It is often easier to distinguish outputs from the internal administrative results if one approaches outputs from the side of the target groups.
Inputs	What the administration is supposed to use to produce outputs. Inputs include such resources as personnel and funding, but also matters coming from the target groups that the agencies take into account or respond to, e.g. a permit application or an investment plan.
Outcomes	The actions expected and the consequences believed to follow the expected actions taken by the target groups in response to the outputs. Outcomes can be further divided into immediate, intermediate and ultimate outcomes. It is clear that hardly any outcome is a result of policy outputs alone, but instead is affected by a variety of other factors as well.

The goals of policies often refer to the last stages in the intervention theory, i.e. outcome 4 "eutrophication reduced" in Figure 2. When the objectives of a policy are very general they can often be specified through the process of reconstructing the intervention theory of the policy. In some cases when the objectives are very general, the process of eliciting intervention theories may reveal explicit or implicit specifications of the objectives. Often intervention theories are useful to make explicit which outcome stage the objectives are intended to address. Objectives are sometimes stated with respect to outcomes earlier in the chain or even in relation to outputs, and there are cases with multiple objectives linked to different stages.

The main use of intervention theories is to guide data collection and analyses in relation to outputs, outcomes and anticipated causal links. The evaluation might focus on the links that are the most crucial for the intervention, the links that are the most uncertain or the links on which it is possible to obtain data. Intervention theories may also assist in interpreting the results, as some assumptions could gain support while others may not and, more specifically, some impacts or lack of impact may be traced to specific links in the intervention theory. It could be that the effects of a policy can be identified but the causal mechanisms are different from those described in the intervention theories. There is a clear opportunity to

learn by reformulating the intervention theories of the policy evaluated and also the theories of related policies. A justification for many evaluations is learning in order to improve the implementation processes. In such cases intervention theories may be crucial in identifying the activities that might be improved.

So far the concept of intervention theories has been generally discussed, but their role can also be discussed in the more specific context of environmental policies (Articles I and III). When evaluating environmental policies using intervention theories, they have a special position. First, they are generally to some degree based on science and scientific concepts. That there are aspects of an intervention theory that can be based on scientific evidence is, partly, a strength compared to situations where scientific knowledge is lacking. However, it is also a big challenge. Handling the science-based part of an intervention theory requires special skills. Science – especially linked to complex environmental and ecological issues – seldom provides one simple and unique view of the matter; and scientific theories may be incorrect and often change over time. Second, due to the long time frames of many, although not all, environmental problems, final outcomes can seldom be evaluated when there is a demand for an evaluation. In these cases the intervention theories can be used to identify which parts of the theory that can be assessed on the basis of scientific theories instead of assessment based on empirical investigations. For example, an intervention theory of a policy instrument used to combat climate change could help in determining the phases the evaluation should focus on and for which phases one would have to use the scientific judgment of, for example, the Intergovernmental Panel on Climate Change (IPCC). (Articles I and III).

Evaluations in general, but intervention theories in particular, can promote learning by individuals, organisations and among different groups and organisations. This is particularly important in the case of environmental policies, since belief systems or views of causal relationships tend to be conflicting. Constructing multiple intervention theories, which make different assumptions on causalities explicit, can foster learning even on its own before any of the assumptions of the intervention theories are empirically examined. (Article I).

4.5. TRIANGULATION

There is no unique and universal answer to how one should empirically examine an assumed link within an intervention theory. For example, to what extent is an investment in a water purification plant a response to a permit limit and to what extent is it explained by other exogenous variables, such as pressures from customers, environmental NGOs or local inhabitants? Often, however, approaches using several types of data and methods instead of only one are appropriate, especially in the environmental policy context. For example, Bartlett (1994, 183) stated: *"Clearly desirable are multiple evaluations, done with a keen appreciation of the strengths and limitations of each approach and a frank recognition of the advantages of others."* Bartlett is by no means unique in advocating multiple methods; often the term "triangulation" is used (e.g. Scriven 1991, 364). Four types of triangulations can be distinguished: multiple methods; multiple sources within one method; multiple analysts; and multiple theories.

For example, the evaluation of the Finnish environmental policy instruments used for the pulp and paper and chemical industries that Articles I, II and IV are based on used statistical analysis in parallel with qualitative analysis of thematic interviews and documents. The data and empirical methods will be described in detail in the next section. The evaluation was, however, an example of all four types of triangulation, with different data ranging from site level information on waste water discharges, air emissions, production and permit conditions to aggregated information for the entire sectors. For this data descriptive statistics, such as means, variances etc. were calculated and examined and a variety of graphical analyses were undertaken. In some cases statistical hypothesis testing was performed, for example for reductions in phosphorus discharges from mills with a phosphorus limit compared to the reduction of those without such a limit (Section 5.1, Article II). A wide range of statistical modelling was performed, including time-series analyses of site as well as sector data and probit and logit models of permit contents (Section 5.2, Article II). The statistical data and analyses were complemented with more than thirty taped and transcribed thematic interviews, which were coded and analysed (Articles II and IV). In addition some cases were chosen for which the background documents generated by the legal processes were examined in detail (Similä 2002). Since all the analyses were undertaken in parallel, the new questions generated by one method were then examined using the other methods, and details required in order to interpret a comment by an interviewee could often be found in the documents or the statistical database.

While one reason for using triangulation is to empirically examine assumed links in intervention theories there are other reasons as well. This brings us back to the concepts discussed earlier. Triangulation is a good approach in order to identify unanticipated effects of policies, which is the key feature of side-effect evaluation. Utilising many data sources and methods is also crucial when multiple criteria are used to assess the policy. Not all data or all methods can have the same role in analysing all the criteria in any multi-criteria evaluation. Nor can they have the same role for all stages in all alternative intervention theories. Emphasising triangulation is rather about starting out from the attempt to define several rather than few, or just one way to address these issues and to stress the interplay between different data, methods and researchers.

Using several methods simultaneously provides many benefits, but it is also challenging and often resource-intensive. Often a multidisciplinary evaluation team is required to ensure the necessary competence. Since all disciplines have their own terminology and style of reasoning, communication between evaluators with their background in different fields is not always easy. A notion such as "binding permit conditions" means one thing to a lawyer and another to an economist. If enough time and opportunities are reserved for interaction throughout the evaluation and all the members of the evaluation team are committed to the idea of triangulation, communication problems can usually be resolved. (Articles I and IV)

4.6. DIFFERENT EVALUATION FOCUSES

The discussion about the concepts of environmental policy evaluation (Section 4.1) started from the different temporal evaluation forms – ex ante versus ex post evaluations – and different evaluands, i.e. focuses of the evaluations. Some key concepts for environmental policy evaluation have now been explored, the relationship between these and the type of policy to be evaluated will be briefly discussed before moving on to the practice of evaluation. While there are numerous potentially interesting evaluands, the discussion will focus on first the difference of using the concepts ex ante and ex post and then on the shift in environmental policies to use economic and information based policy instruments alongside regulations (Section 2.2).

The proposed concepts were previously largely discussed with ex post evaluations in mind; while they all can also be used in ex ante evaluations, their role becomes different. Starting with side-effect evaluation, it is obvious that truly unintended side-effects cannot be

evaluated in advance. However, taking side-effects into account when designing policies and undertaking pre-assessments can clearly broaden the scope of the anticipation of side-effects. Furthermore, debates before a policy is adopted often concentrate exclusively on main effects and few side-effects are discussed. An ex ante evaluation can thus widen the span of attention. For example, the environmental impact assessment conducted during the preparation of the National Climate Strategy made information available on several environmental side-effects such as air emissions causing acidification (Hildén et al. 2001). As for the criteria utilised effectiveness and efficiency tend to dominate policy formation. The multi-criteria approach advocated also forms a wider base for normative discussions of the options in advance.

Assumptions about how policy interventions will influence different actors and their behaviour are central when new policies are debated. Therefore, intervention theories may be used both ex ante and ex post. Triangulation, starting out from several rather than just one way to tackle these assessment challenges and to stress the interplay between different data, methods and researchers is surely also relevant for ex ante evaluations. Data availability might be a more severe limitation for some methods before policies are adopted but not for the idea of triangulation as such.

As previously noted, environmental polices in Finland as well as in other countries, have expanded from almost exclusively regulation to a more frequent use of economic and information based instruments (Section 2.2) The question therefore arises whether there are differences when utilising the evaluation concepts for different policy instruments. First, the concepts can all be applied to all kinds of policy instruments (the practice of which will be discussed in Section 5.5). Second, there are, however, some clear distinctions affecting the applicability of the concepts to economic and information based instruments.

In the case of regulations they are generally implemented and enforced through public agencies and information tends to be quite freely available, being especially readily accessible for Finnish environmental permits (Hildén et al. 2003). Some economic instruments, especially subsidies, tend to represent the other end of the spectrum; detailed information is generally confidential, based on the need to protect commercial secrets. Information based measures are often implemented in collaboration with private companies, such as auditing firms in the case of environmental management systems and advertising agencies in the case of information campaigns. Transparency is therefore often restricted since the private partners can gain competitive advantage from exclusive knowledge. Limited access to information becomes a problem if access to information has not been fully considered and taken into

account when contracts have been signed. While access to information affects how extensively triangulation can be used, it also stresses the role of transparency as an evaluation criterion.

Justifications for economic policy instruments are frequently based on economics. Read more thoroughly, the economics literature, however, also provides a lot of assumptions, limitations and criticism of these instruments. Economics is thus a particularly important source while constructing intervention theories for economic policy instruments. For example, it has often been argued that economic instruments are superior for promoting innovations, since they impose a cost on emissions irrespective of their level and, therefore, provide a continuous incentive to innovate (e.g. Jung et al. 1996). This claim, however, has also been criticised, for example for being based on the unrealistic assumptions that it is politically feasible to set environmental taxes at high enough levels and without exemptions (e.g. Kemp 1997). Since economic instruments are generally promoted by recourse to arguments of efficiency and cost-effectiveness, these criteria become particularly important. Economic data, calculations and methods are in an especially important position because they are a requirement for utilising these criteria. A broader set of criteria could still be illuminating also in the evaluation of economic instruments.

Such criteria as transparency, participatory rights and acceptability are especially important when evaluating information-based environmental policies. These policies generally try to influence the knowledge as well as the attitudes of the target groups, and unless such action is widely accepted, it cannot be justified. The control of attempts to influence attitudes through public policies requires both transparency and participation. As discussed by Vedung and van der Doelen (1998, 109) information-based instruments are built on paternalism, i.e. *"the government intervenes in order to effect the recipients to act in their own not very well-understood interests"*. Paternalism is inherent in all policy instruments. Information-based policies, however, deal with the foundations of democracy – preferences and attitudes of individuals and groups as a basis for ideas, deliberation, and choice – therefore paternalism is especially relevant for these instruments, reinforcing the need to use the criteria transparency, participatory rights and acceptability.

Considering the expansion of environmental policy in scope as well as in form, it would not be wise to use the same amount of resources to evaluate all policies. The question what policies should be evaluated then arises. In 2005 a cross-sectoral working group suggested improvements for ex ante impact assessments of law drafting in Finland (Finnish Government Working Group for Planning and Steering of the Development of Law Drafting 2005, 219).

Their suggestion was based on a screening of proposals at an early stage; categorising them into broad, normal and limited ones based on their anticipated effects; and determining the scope of the impact assessments based on the categorisation. This is clearly also a useful approach when focusing limited resources for ex post evaluations. More generally the focus should be on evaluations expected to provide the greatest opportunities for learning and development or those most crucial for accountability. When selecting these evaluations the scale of the anticipated effects must be considered, but so also must the windows of opportunity for changes and the ongoing policy discussions.

5. PRACTICAL EXPERIENCES OF ENVIRONMENTAL POLICY EVALUATION

5.1. EMPIRICAL MATERIAL

Two of the Articles (II and IV) on which this thesis is based are largely empirical, while the three others use empirical material mainly to exemplify the conceptual arguments. All the articles, to some extent, make use of the same empirical material. That is the material gathered and used in order to evaluate the Finnish policy instruments used to regulate the pulp and paper and partly chemical industries (Hildén et al. 2002). Although the overall evaluation also covered energy taxation and environmental management systems, only the waste water permits used for pulp and paper mills and the material and methods used in their evaluation will be presented, since they are the focus of this thesis and the articles only cover this part.

Until March 1st 2000, when the new Finnish Environmental Protection Act came into force, water permits were issued on the basis of the Water Act, which came into force in 1962. Finnish waste water permits were issued case-by-case, based on an application. A special – although not unique – feature internationally as well as compared to the air permit system, is that there were no standards to explicitly guide the individual permit decisions, such as general emission limit values or binding environmental quality objectives. The water permits were issued by three Water Courts. Decisions by the Water Courts could be appealed to the Water Court of Appeal or to the Supreme Administrative Court (Hildén et al. 2002, Articles II and IV).

Every permit decision was preceded by a process. Although this process formally started when an application was submitted by the mill, there were usually many contacts between the mill and the environmental authorities already before that. After an application had been submitted the process was often quite slow; as long as 16 years could pass between the date of submitting an application and the issuing of the permit decision. Although these decisions were made in the 1970s, some permits issued in the 1990s were also based on applications made more than 10 years earlier.[5] The processes became even longer if any party appealed.

5 The mills were able to start or continue operating during the permit process.

The quantitative mill level data is based on information about waste water discharges, production and permit conditions. The data on discharges was obtained from the official database of the environmental administration. The information on permit conditions consists of the quantitative limit history of all waste water permits for the pulp and paper mills operating in 1998. The information was collected by starting with the permits in force in 1998 and working backwards based on the reference of which permit the more recent decision replaced. Only such decisions where at least one limit value was set or changed were included. The water quality from the nearest observation point of each mill was used and the state of the waters was calculated for time periods of 5 years. (Article II)

There are 192 decisions on present or past permit limits for the mills operating in 1998. Of these, four decisions concerned several mills, with separate sections determining the limit values for the discharges separately for each mill; these are treated as separate decisions. Even though they are part of the same legal decision, they were decided at the same time and through a common process, the conditions are separate. Taking this into account the data is based on 213 decisions, of which 166 were made by the water courts and 47 by the courts of appeal. (Article II)

During the spring of 2000, thematic interviews were carried out with company representatives and authorities. Nine employees of the pulp and paper industry were interviewed, representing all three major companies. Of these, six worked at mills in different parts of Finland while three were working at company headquarters. Seventeen employees of the authorities were interviewed, six from municipalities, six from regional agencies and five from the national authorities. All the interviews were recorded, transcribed and then coded. Most of the interviews were between one and a half and two hours long. (Article II)

Every expert contacted agreed to participate in the interviews for the evaluation. The thematic interviews provided in-depth information on how the environmental impacts of the mills had developed, what actions had been taken and when, as well as perceptions regarding why and what role different factors played, including the permit limits. The goal of qualitative analyses is not to produce data that can be generalised through statistical procedures or otherwise become "universal" (e.g. Yin 1994). The interviewees were, therefore, not chosen through any sampling techniques, but as people with a lot of experience of environmental permitting and environmental technologies of pulp and paper production, with different backgrounds. (Article II)

The quality of data collected by interviews depends on the interviewers and their skills. In this evaluation several different individuals collected and analysed the material which reduces systematic errors. In the interviews the same questions were discussed from many angles in different contexts and during different phases of the interviews. The different statements were then crosschecked with the help of detailed coding in the computer programme N-Vivo. In addition written documents, for example the background information of the permits, were used to confirm statements whenever feasible. (Article II)

A third type of material used in the evaluation is the material produced for the legal decision-making. This material was used in the evaluation (Hildén et al. 2002) and by Similä (2002) but was also used as a secondary source in Article IV. The document, which is an output of the decision-making, contains two parts: the permit itself, i.e. the decisive part of the document and a report. In the report the history of the procedure is described, starting from the application. The documents also include an overview of the plan of the environmental measures to be taken as well as opinions and arguments of different actors concerning the key conditions of a permit. These documents are huge, some of them several hundred pages, it was therefore not possible to analyse all these documents for all the permits. Similä (2002) examined the whole permitting histories of six pulp and paper mills in detail. The total number of permits including an emission limit value for these six mills was 23, i.e. three or four per mill. The documents were analysed qualitatively in order on the one hand to identify the typical features of the actual permits granted and on the other hand to gather information from the documents on the impacts of the earlier permits. In many cases information from the permit documents could be directly combined with information from the interviews.

5.2. FINDING IMPACTS THROUGH TRIANGULATION

It was well known long before the evaluation that the discharges of many waste water pollutants had dropped significantly, although the pulp and paper production had increased. In other words the discharges had been decoupled from production (Figure 3). A similar development could also be observed for some air emissions. The open issue, however, was to what degree these developments could be attributed to the environmental policy instruments and their specific features.

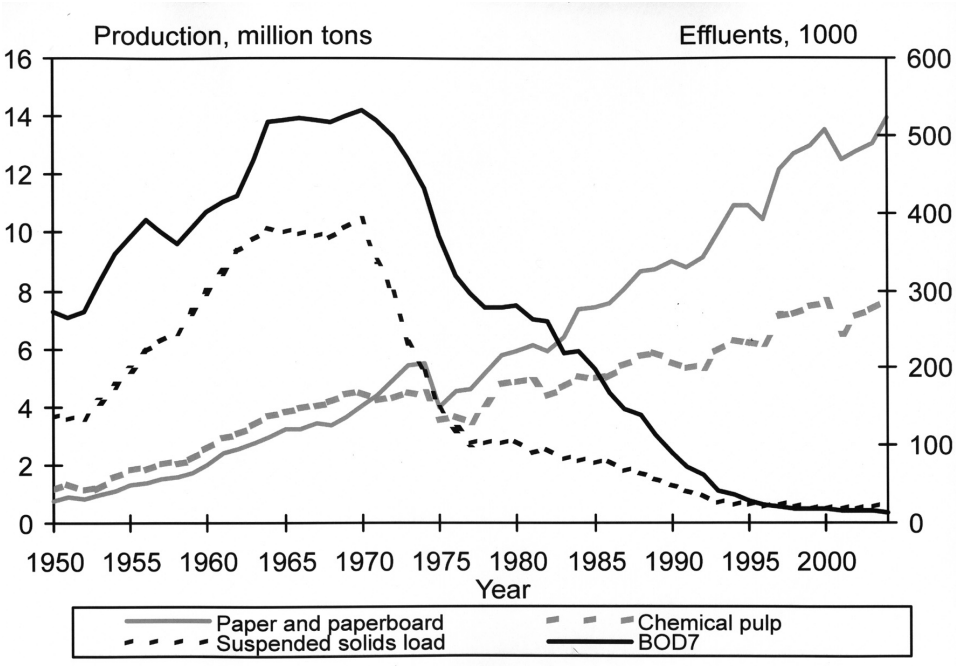

Figure 3. Decoupling of production and some water discharges in the Finnish pulp and paper industry (Source: The Finnish Forest Industries Federation and the Finnish Environmental Administration)

All pulp and paper mills listed by the Finnish Forest Industries Federation in 1998 were operating on the basis of some water permit. Some mills have had permits with quantitative limits since the beginning of the 1970s (Figure 4), whereas one mill obtained its first limit as late as 1995 although it had been operating since 1965. Some mills have had the same limits for many years, whereas there are also mills that have operated based on limits that have changed rather frequently. Some individual permits have included several limits for different time periods for one type of discharges, e.g. BOD_7 (Biochemical Oxygen Demand, measured over 7 days). While the mill with the greatest number of different limit values on the total BOD discharges had had 8 different limit values, one mill had had only one limit for its total BOD discharges, which came into force in 1992. Two of the mills had had only two different limits even though they had had limits since the 1970s. (Article II)

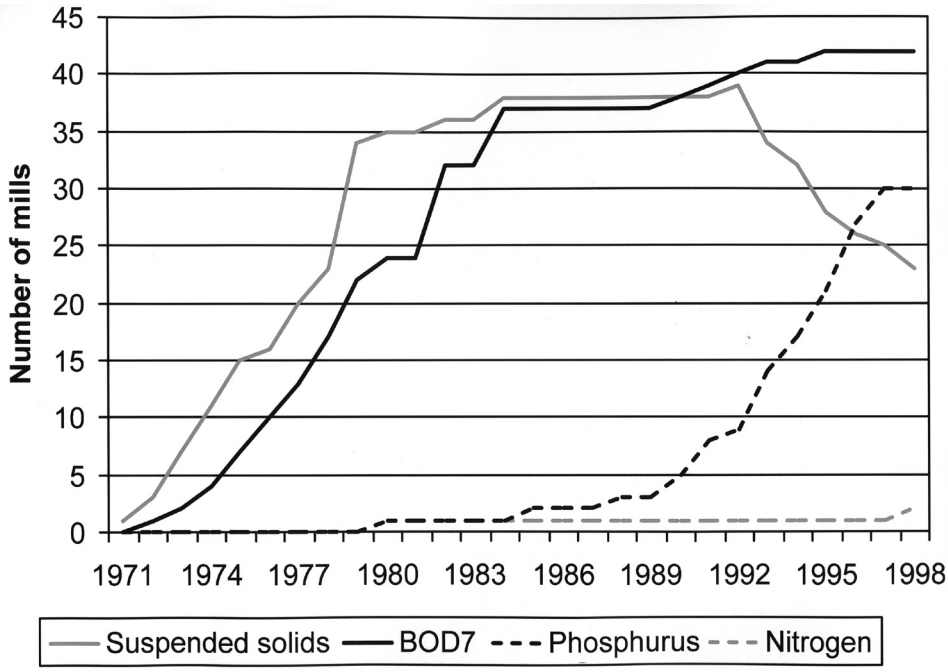

Figure 4. Limits on total discharges included in permits of the Finnish pulp and paper mills operating in 1998 (Article II)

Absolute limit values generally become tighter when they are changed. For example, in the case of the BOD limits there are only 9 cases in which the absolute BOD limit has been raised. This corresponds to about five per cent of the limits. Of these, four have been temporarily increased, i.e. the decision has stipulated a later return to the same or a lower level than before. The temporary increases have been issued, for example, during the period when an activated sludge treatment plant was being built and started up. (Article II)

Many different types of statistical analyses were undertaken in order to examine the impact of BOD and phosphorus limits on the discharges, these are reported in detail in Article II. The choice of analyses depended on how many mills had a limit value and for how long these limit values had been issued (Figure 4). It was thus possible to use time series analyses to examine the effects of BOD limits on discharges at the plant level but not for phosphorus. Instead the temporal development of absolute as well as relative phosphorus discharges of the mills with permit limits was compared with that of the mills without any limits and the significance of the differences was tested. (Article II)

Based on the mill-specific time series models, a few cases where the BOD limit had a statistically significant and positive coefficient were found. In some cases the models actually explain the data fairly well. For example, the differential model of mill 3 has an R^2 of 0.91, the residuals are normally distributed etc. (Article II) The statistical analyses required further analyses since most of the time series were found to be non-stationary. There were still some models clearly indicating the effects of the limits for the specific mill. The mills for which the time series results are clear are rather few, which is no surprise considering the rather short time series. (Article II)

In 1998 all mills had considerably lower phosphorus discharges compared to their maximum discharges. However, there is no statistically significant difference between the mills with a limit on their total phosphorus discharges (on average a 76.6% reduction) and those without it (on average a 70.8% reduction). Although the reductions in the phosphorus discharges in relation to the production of the mill are on average greater than the reductions in absolute discharges the general result holds, there is no clear difference between the mills with a limit on their total phosphorus discharges (on average a 88.0% reduction) and those without a limit (on average a 84.4% reduction).

Another statistical method used to examine the impacts of the limits on phosphorus was to estimate the trends of the discharges for every mill and then the possible difference between the groups was tested. A clear difference between the slope for the mills with a phosphorus limit and those without was found (a slope coefficient of -803.9 vs. one of -81.0). Since the analyses of the inclusion of phosphorus limits revealed that phosphorus limits were introduced earlier for mills with high phosphorus discharges, this could explain the on average faster reductions. (Article II)

Not only were analyses at mill level undertaken; some modelling of the aggregated pulp and paper sector was also carried out. It has been claimed that when phosphorus limits started to become common, those mills which did not yet have a phosphorus limit in their permits also took action in order to reduce their phosphorus discharges. By estimating some models of the total phosphorus discharges this argument was examined. The results indicated that the phosphorus discharges of the sector were reduced when the number of mills with phosphorus limits increased. The results should be interpreted with caution, however, partly because the coefficient is not significantly different from zero (there is a 17 per cent probability that the coefficient is zero), and since recursive estimations show that the parameter values are heavily dependent on the sample length. (Article II)

As a summary of the statistical analyses of BOD and phosphorus one can conclude that clear effects of permit limits on the discharges of many but not all mills were found. The fact that some mills got limit values very late, while others had limits that did not change for decades, reduced the impacts. (Article II)

Following a decision by the Ministry of the Environment in June 1989, permits for pulp mills started to regulate chlorine bleaching through specification of the AOX limit values. The decision by the Ministry stipulated annual mean chlorine discharges less than 1.4 kg per ton of bleached pulp produced, determined with the AOX method as a target (Ministry of the Environment 1989b). In 1997 all mills had AOX limits included in their permits. All the limits, however, were set clearly higher than the annual average discharges of the mill concerned during the year of the decision. Despite this, 1 year and 4 months on average elapsed before the limits came into force. When all mills had AOX limits in force their average annual AOX discharges per ton of bleached pulp was only about 0.2 kg. The reduction of the discharges was thus much faster than the Ministry target. (Article II)

All interviewees were able to give some examples of reduced discharges that in their opinion were attributable to the permits. The interviewees were also able to give specific examples where a water permit or a particular limit had caused a company to take actions it would not otherwise have taken. However, examples have also evinced where discharges had declined but due to other factors than permits. The greater know-how than before in several areas such as law and engineering was seen as a clear example of an impact outside the target area caused by the information demands created by the permit system. (Articles II and IV)

In the 1990s new bleaching processes for pulp were introduced in order to decrease the discharges of chlorine compounds. All interviewees from the mills, however, used this as an example of new processes spreading mainly because of customer demand. They thus argued that the actual limit values in the permits played little or no part. This view was contradicted by some, but not all interviewees, who worked for the environmental administration. They emphasized the role of the ministerial decision of June 1989. (Article II)

Based on the detailed legal documents he analysed Similä (2002, 154) was able to conclude that four out of the five mills studied had once been forced by some permit conditions to adopt a new major end-of-pipe solution. By combining the information with interviewee information Similä (2002) could also conclude that the fifth mill had once been forced to adopt a more advanced end-of-pipe solution than originally planned. In addition there were also five cases that resulted in smaller actions. Although several clear effects were found Similä

(2002) also showed that not every permit or every limit value had an impact.

The impacts of permitting cannot be evaluated by merely comparing permit limits and discharges as the results of the different methods showed. Even if the permit limits were non-limiting at any given moment, reductions in discharges could still have been caused by the permitting process. Often new technological solutions were adopted before the limits entered into force or sometimes even during the negotiations. There may not therefore afterwards be any limiting permit limits. The advantages of triangulation in order to evaluate impacts were clearly demonstrated.

5.3. EXPERIENCES OF USING OTHER EVALUATION CRITERIA THAN IMPACT

In addition to the impact criterion already discussed, Article IV summarises the experiences of utilising other evaluation criteria and relates them to the different materials and methods used. Table 4 summarises the role of the main types of methods for the criteria.

Table 4. The role of different methods when using other criteria than impact in the evaluation of Finnish permits (based on Articles IV and II)

	Method		
Criteria	Statistical	Qualitative	Reconstruction of processes (Similä 2002)
Relevance	Used	Central	Used
Effectiveness	Statistics central	Used	
Efficiency	Used	Used	Used
Flexibility	Central	Central	Central
Predictability	Modelling central	Central	
Persistence	Used	Central	
Acceptability		Central	Used
Transparency		Central	Central
Equity	Used	Central	

Not all the criteria, nor all aspects of how the methods were used will be discussed here (see Article IV for a more extensive discussion). Instead, some examples demonstrating the usefulness of using more criteria than just impact and examples of the interplay between the methods will be discussed.

The effectiveness of the permit systems was mainly good in the sense that targets were generally reached, and this was at least partly due to the permits, but also affected by other factors such as technological change. For example, the quantitative targets of the water protection programmes were generally met. The goals for reductions of SO_2 air emissions were also reached, but some recent NO_x and VOC goals were not achieved. A question raised was whether the observed effectiveness said more about the policies or the level of ambition when setting the goals. A key civil servant from the Ministry of the Environment has stated that: *"In Finland the politically determined goals for environmental administration and environmental protection have generally been realistic or actually cautious. Accordingly, achieving them has seldom imposed big problems."* (Ojala 1997, 75) (Article IV)

The results of the evaluation confirmed the expected flexibility of the Finnish permit systems and also showed that the regulated companies as well as the authorities valued the flexibility of the systems. Flexibility has implied the option to differentiate the requirements based on such factors as local conditions and the features of the mill concerned. The evaluation showed that there were wide variations in when specific limits had been included in the water permits of different mills; in the strictness of these limits; in how often they were renewed; and in the times between decisions and when the limits entered into force. The praxis that the limits in the water permits could be temporally relaxed when new abatement technologies were being tested or introduced is also an example of the flexibility that has improved the dynamic efficiency of the permit systems. A normative flexibility was demonstrated, i.e. tightening of emission limit values, and introduction of parameters occurred without any change in the regulations. (Article IV and Similä 2002)

A flexible permit system, especially one without standards could become very uncertain for all parties, predictability thus becomes an important criterion. Since about half of the 56 water permits issued during the years 1987 to 1997 included a phosphorus limit while the others did not, it was possible to use modelling to examine this aspect of predictability. The factors affecting the introduction of limit values for phosphorus in the water permits were tested with logit and probit models (Article II). The results showed that the introduction was not random, instead it was largely predictable based on some characteristics related to the mill. Mills with high phosphorus discharges during the year preceding the permit decision and mills situated in the jurisdictional area of the North Finland Water Court had an increased likelihood of getting a phosphorus limit in their permit, whereas those mills situated along the coast were less likely to get a phosphorus limit in their permits compared to other mills. The

informal preparatory phase and the often long formal process ensured that the permit content was seldom a surprise, especially not for the company. (Article IV)

The evaluation showed that the Finnish permit systems were fairly transparent and that the public provided input to the decision-making in water and air pollution issues. The unrestricted access to permit decisions, partly to their background material and the availability of factory-specific emission data has become so natural in Finland that it is taken for granted and not questioned by any party. The openness has improved the predictability of the system and made possible something loosely referred to as "the general policy" (Similä 2002) even without standards. The evaluation demonstrated great differences in resources and networks between actors. For example, mills generally interacted a lot with consultants and technology providers, while local authorities interacted more with each other. In addition, the informal phase before an application has been made is not that transparent, simply due to its informal nature. This partly reduces the "formal" transparency. (Article IV)

5.4. FINDING SIDE-EFFECTS AND UTILISING INTERVENTION THEORIES

The practice of utilising two of the evaluation concepts emphasised in Section 4 has been discussed extensively so far, that is multi-criteria evaluation and triangulation. Something on the practice of side-effect evaluation and on actually utilising intervention theories should also be said. The fact that the practice of actually using these concepts in evaluations is not so extensively discussed in the articles of this thesis does not imply that they are less important.

A side-effect of the permitting systems was an increased demand for education and general know-how in the environmental field. This side-effect has turned out to be essential both with respect to a sustained improved state of the environment in the surrounding areas of point source polluters and with respect to the ability of the environmental administration to gradually expand its scope of environmental policies. The scope has expanded from water to air and waste and then towards more integrated and catch-all approaches. (Article IV)

The high degree of transparency of the permit systems has assisted public discussions about emissions, and even individual plants have had to justify their environmental performance on the markets. Although the transparency as such has been intended, this effect can be seen as a largely unanticipated side-effect of transparency. (Article IV)

The temporary relaxations of permit conditions already referred to were part of the intended flexibility of the system. Nevertheless the crucial role of this feature for the development and diffusion of new technologies was an unanticipated side-effect. (Hildén et al. 2002, Mickwitz et al. 2005)

Several forms of technical development that are not policy induced have taken place. These include the abolishing of sulphite pulp production mainly due to economic reasons and the decreased water content of sludge in order to increase energy efficiency. The abolishing of sulphite pulp production reduced BOD discharges as well as SO_2 emissions as side-effects. Burning dryer sludge not only increased energy efficiency but also reduced the SO_2 emissions. Since these technical developments were not caused by the policy it would be too much to label the side-effects of the technical development side-effects of the policies. But when considering the benefits resulting from the fact that the Finnish permit systems have not hampered technological change, these side-effects should also be included.

Intervention theories in the form of more detailed versions of that presented in Figure 2, were used as tools when planning and evaluating the links between permits and discharges. In addition special intervention theories were developed for some features of the permits. An intervention theory of the underlying assumptions of the research and development (R&D) requirements included in some, but not all, permits will be used to exemplify the practice of using intervention theories (Figure 5).

The intervention theory in Figure 5 can be related to a case described by Similä (2002, 156). In this case a mill had an R&D requirement in its permit to study the potential for reducing the phosphorus discharges below a particular limit. A decade later such a limit was included when a new permit was issued. Although a new technology was adopted and thus the phosphorus discharges were reduced, it was not in this case because of the specific R&D effort undertaken. The results of the R&D had indicated – wrongly as is known in hindsight – that it would be impossible to reduce the phosphorus discharge to the proposed level. The intervention theory in Figure 5 constitutes one version of the many steps between the R&D requirement and the reduced discharges. It also shows that when new R&D findings lead to additional costs due to new investments the incentives for a firm to ensure that such new results are produced are not that great.

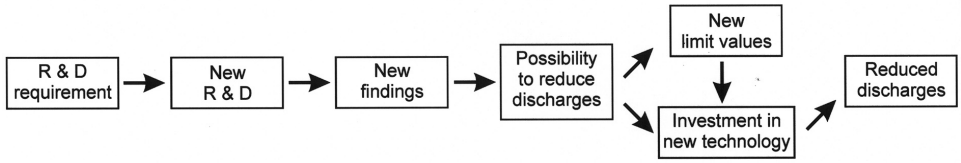

Figure 5. An intervention theory of the research and development requirements of Finnish water permits (Mickwitz 2000)

5.5. EXPERIENCES OF THE CONCEPTS IN OTHER CONTEXTS

So far the practice of environmental policy evaluation based on the proposed concepts has been discussed only through the experiences from evaluating permit limits for pulp and paper mills. Although many insights are provided, it is justified to ask how these experiences relate to the other policy instruments and sectors, the more so since many more policy instruments exist and many more targets than large-scale industrial point source pollution (Section 2.2). The practice of utilising the concepts will be briefly discussed first in light of experiences of utilising them for the evaluation of economic and information-based instruments for pulp and paper mills (Hildén et al. 2002) and then in the light of the experiences of evaluating the two first years of permitting based on the new Environmental Protection Act for all sectors regulated by it (Hildén et al. 2003).

The effects of energy taxation on the energy use of pulp and paper industry and emissions were found to be slight. This was mainly due to a lower tax level for industry, an exemption from the tax for electricity produced with wood or wood based fuels and a tax rebate to energy intensive companies. However, since a major justification for the detailed taxation system has been not to impair the competitiveness of Finnish export industries, this aim has been achieved. (Hildén et al. 2002, 112) It is likely that the existence of the energy taxation and the almost continuous debate about it as a side-effect has provided signals promoting voluntary energy agreements (Hildén et al. 2002, 72).

Adopting environmental management systems, especially EMAS, has become very popular among pulp and paper mills. The evaluation conducted when most mills had used the systems for only a short time suggested that examples of specific improvements could be found. The triggers of these environmental improvements were generally the systematic screening of the activities, the consolidation of internal environmental awareness or the systematic handling

of complaints and other external signals. Examples of such improvements concerned accidental air emissions and noise disturbing neighbours. Specific improvements attributed to environmental management systems were generally either low cost or even resulted in cost savings. (Hildén et al. 2002, 75-81)

The new Environmental Protection Act (86/2000) came into force in March 2000, and already the same spring the Ministry of the Environment launched an exceptionally broad evaluation effort. One part of this effort was a research based evaluation, conducted as a follow-up to the frequently mentioned evaluation of policy instruments for pulp and paper mills (Hildén et al. 2003). The other part was an assessment produced by a working group led by the ministry but with a broad participation from interest organisations (Working group for assessing the implementation of the Environmental Protection Act 2003). There was a many-sided interaction between the two efforts.

Intervention theories based on documents were used to establish the main intended effects of the new permit system compared to the old ones (Similä and Hildén 2003). The main effects the reform was supposed to ensure were: that one permit system should be cheaper than several parallel systems; that an integrated system was supposed to be more efficient in achieving environmental objectives, since problems would not be transferred from one area to another, because holes between systems would be filled and prioritising would become possible; and that public participation would be increased (Similä and Hildén 2003, 16). While a main driver of the reform was to transpose the Integrated Pollution Prevention and Control Directive (96/61), it was domestically decided to extend the scope – by including all activities that had earlier been required to have any permit – from the large installations covered by the Directive to much smaller units.[6]

When such a major reform is evaluated so recently after coming into force care has to be taken that the long-term benefits are not underestimated and the relative role of initial administrative problems is not over-emphasised. Nevertheless some relatively firm conclusions can be drawn. The number of permit decisions became very large, partly due to stipulations in the act and partly due to external factors. One such factor was the structural change in Finnish agriculture, reducing the number of farms but increasing the number of big farms required to have a permit. The backlog of permit applications – expected to continue for several years

6 There were 632 installations covered by the Directive in 2002 (Mickwitz et al. 2003, 35), while the total number of sites was expected to be between 25 000 and 30 000 (Sjöblom et al. 2003, 83).

– has resulted in growing administrative costs, especially at the regional level. The anticipated cost-reductions were overestimated because a vast majority of the sites obliged to obtain a new environmental permit had only had one permit before the reform. There was therefore no synergy from the integrated permit for these installations. While some installations will save costs in the long-run, because of the integration, the application requirements became somewhat more detailed for most applicants, requiring e.g. energy efficiency analyses. (Hildén et al. 2003, 184-6)

During the first two years few environmental gains had been achieved through the integration of all environmental aspects into a single permit. Very little prioritisation of different environmental effects was reflected in the limit values or other stipulations of the permits. It seems likely that a longer learning process is required before the full effects of permit integration can be realized.[7] The main environmental benefits during the first two years occurred through the possibility to deal with new aspects of such non-industrial activities that previously had only one permit. For example, restrictions on waste treatment in fish farming could not be issued in the old water permits, but can now be included in the integrated environmental permits. The reform has somewhat increased the participation opportunities, by extending the rights to participate from those directly affected to encompass environmental non-governmental organisations. This right, however, was seldom used during the first two years. The reform also made the participation opportunities clearer. (Hildén et al. 2003, 185-7)

The multi-criteria discussed in Section 4.3 were also used to evaluate the Environmental Protection Act. An interesting result, especially in relation to the earlier obtained results for pulp and paper mills, was the assessment of the flexibility of the new law as well as its implementation. The Act provides much flexibility with respect to content; the case-by-case judgment of individual limits is even greater than before. There is also regulatory flexibility in the sense that permits can be complemented with sector specific norms, which surely would then impair the flexibility of individual permits in the sectors concerned. The only flexibility with respect to the backlog of permit applications has been to employ more civil servants (Hildén et al. 2003, 190). The evaluation defined a clear risk that from an environmental

7 After the Water Act came into force in 1962, it also took a long time before quantitative limits became included in the permits and permitting processes were often long in the beginning (Sections 5.1 and 5.2).

point of view too many resources would be used for permitting and too little for other tasks, such as enforcement. It was subsequently suggested that the possibilities to substitute case-by-case permits with general norms would be seriously investigated for small-scale relatively homogeneous activities, such as petrol stations or stone crushing plants. While such a re-regulation would seem well justified in light of some criteria, it would also reduce the possibilities for public participation and the flexibility to take local conditions into account. (Hildén et al. 2003, 193-5) Although flexible case-by-case permits have many advantages while regulating pollution from large-scale point sources such as pulp and paper mills, they are clearly no panacea; their value depends both on the context and the evaluation criteria.

6. EVALUATING THE EFFORTS TO EXPAND ENVIRONMENTAL POLICY THROUGH POLICY INTEGRATION

6.1. POLICY INTEGRATION AND THE FOCUS OF ITS EVALUATION

Pollution and other harmful environmental effects often follow from activities greatly affected by public policies. Examples of such activities are transport, agriculture and energy production. Because these activities are extensively influenced by public policies, many have argued that environmental harm could be reduced if environmental aims were taken into account when these policies are designed and implemented. The harmful side-effects of these policies could thus be anticipated and reduced. (e.g. Weale 2002, 203) If we would stick to viewing environmental policies based on the purpose of a policy (Section 2.1) this would imply expanding the scope of environmental policy, since if, for example, transport policies also have environmental aims these policies would be included among the policies intended to affect the environment. Usually the term environmental policy integration is used for the inclusion of environmental policy aims into other policies. (Article V)

Based on the definition of policy integration by Underdal (1980, 162), Lafferty and Howden (2003, 9) define environmental policy integration as:

> *"the incorporation of environmental objectives into all stages of policy-making in non-environmental policy sectors, with a specific recognition of this goal as a guiding principle for the planning and execution of policy;*
>
> *accompanied by an attempt to aggregate presumed environmental consequences into an overall evaluation of policy, and a commitment to minimise contradictions between environmental and sectoral policies by giving principled priority to the former over the latter."*

Political commitment to environmental policy integration can be found in Chapter 8 of Agenda 21 adopted in Rio 1992, which calls for the integration of environment and development at the policy, planning and management levels (United Nations 1994). The EU has also made commitments to environmental policy integration. Starting with the EC's Third Environmental Action Programme (1983) and reinforced in the following action programmes, the need to integrate environmental considerations into the formulation and

implementation of all sectoral policies has become a key policy principle (Liberatore 1997, 108). Already before Rio, in 1987, the Single European Act, in Article 130r(2) stated that *"environmental protection requirements shall be a component of the Community's other policies."* The role of environmental policy integration was further strengthened in EC legislation when the Treaty of Amsterdam, in 1997, in Article 6 stated: *"Environmental protection requirements must be integrated into the definition and implementation of the Community policies..."*.

These political demands to integrate environmental policy concerns into other policies, however, are not unique. There are also several other calls for policy integration, e.g. related to gender issues, regional development, entrepreneurship and employment. The terms "horizontal priorities" or "mainstreaming" are often used when these demands are articulated. The wishes for policy integration can, at least partly, be seen as consequences of the development of government. Peters (1998) has pointed out that policy co-ordination has turned out to be more difficult because: the role of government has expanded, new agencies have been formed, decisions have been transferred from the ministries, participation of clients as well as employers has increased, and many policy issues have become broader and more complex. In Finland, too, public sector reforms have made co-ordination of the governmental policies more challenging (Temmes 1998). Three external experts stated that *"Possibly the most vital task facing political and administrative authorities in Finland is creating a vigorous new culture of working across existing structures within the political system in order to confront more effectively the complexities of contemporary policy problems."* (Bouckaert et al. 2000, 17) The requests for policy integration can be viewed as one of several responses to the lack of co-ordination. (Article V)

Policy integration should be evaluated in order to learn from the efforts undertaken and to develop the practices of policy integration. Another important reason is to examine the sincerity of the commitments and efforts to integrate policies. While policy integration can be an effective way to coordinate policies, it is also an old way to divert attention and to obliquely resist the political goals one claims to support in the integration statements. (Article V)

Policy integration could, in principle, take place in many phases of the policy process. Assuming that there is a general political commitment that policies should be integrated, this should then be reflected in sectoral policy strategies, as well as in the instruments through which these strategies are implemented. The suggested evaluation approaches have so far mostly been process focused, for example, OECD (2002), Lafferty and Hovden (2003) and Jacob and Volkery (2003). In contrast to this focus, Article V calls attention to the primary

idea of policy integration; it is not just to change bureaucracies but to actually change the real world. The justification for integrating environmental aims into transport policies is the perception that the environmentally harmful side-effects of mobility could thus be reduced. Article V argues that the perspective by Lafferty and Hovden (2003, 12) *"focusing primarily on process and policy; not on the actual consequences and effects of governmental initiatives"* should be extended and even shifted to the actual consequences (outcomes in Table 3). While an evaluation perspective focusing on outcomes is very demanding, it cannot be avoided if the intention is to learn and to develop policy integration into an efficient way of achieving environmental policy aims. (Article V)

The importance of including outcomes in an evaluation of policy integration can be illustrated by the example used in the previous discussion about an intervention theory of R&D requirements in permit decisions (Section 5.4). If only the process or the output were evaluated the mill required to examine the possibility of reducing its phosphorus discharges would be an example of successfully integrating innovation policy aims into environmental policy. Based on this the proponents of innovation policy integration could require more similar permit conditions. Only after it is demonstrated that the results of the R&D (i.e. an outcome) showed that it was not possible to reduce phosphorus discharges, and that subsequently history proved this finding wrong (another outcome) can a practical discussion about integrating innovation policy aims into environmental policy start.

6.2. EVALUATING ENVIRONMENTAL POLICY INTEGRATION INTO TECHNOLOGY POLICIES

Evaluating environmental policy integration at all stages from strategies to outcomes can be exemplified in the field of technology policy (Article V, Kivimaa and Mickwitz 2005). In this context environmental aims are considered as aims to improve the quality or to prevent the deterioration of the natural environment, and technology policy is considered to be policies planned and implemented by some specific institutions (Section 2.1).

The Science and Technology Policy Council, an advisory body chaired by the Prime Minister, coordinates science and technology policy in Finland and issues strategic reviews. The reviews of the 1990s emphasised the importance of environmental technologies, sectoral research and co-operation between sectors such as environment and technology (Science and

Technology Policy Council 1990, 1993, 1996, 2000). The focus of the latest review (Science and Technology Policy Council 2003), however, was on internationalisation, with hardly any mention of environmental issues. (Article V)

The Finnish National Technology Agency, Tekes, is the central institution for the implementation of research and development (R&D) policies for new technologies in Finland. In 2004, Tekes's technology strategy was based on eight thematic areas, one of them, sustainable development including e.g. environmental technologies and life cycle solutions (Tekes 2004a). Since all themes should complement and overlap each other, sustainable development can be seen as a part of the whole strategy. Therefore direct effects on social, environmental and welfare aspects should be used as criteria when decisions on which projects to fund are made (Tekes 2004a, 2004b). (Article V)

The main policy instrument through which technology policies are implemented is technology programmes. In Finland technology programmes were introduced in the early 1980s, and in 2003, 180 million Euros, i.e. 46 per cent of Tekes funding, was allocated through these programmes (Tekes 2004b).

In Article V eleven technology programmes carried out between 1988 and 2002, were studied to evaluate environmental policy integration at the policy instrument level. The programmes chosen all affected the pulp and paper industry in order to utilise the general understanding of the sector developed through earlier research (Articles II, IV and Kivimaa and Mickwitz 2004). Some of the technology programmes were limited to the pulp and paper sector, e.g. RAINA, while others covered wider areas but also included some research related to pulp and paper production, e.g. LIEKKI. (Article V)

The degree of environmental policy integration in the technology programmes evaluated was found to be issue specific. In programmes focusing on environmental technologies principled priority for environmental issues could clearly be observed. The diminished role of environmental issues in the most recent strategy, however, is also reflected on the instrument level. The focus of the pulp and paper sector programmes has consequently shifted from environmental issues to quality concerns. Yet Tekes still have some large research programmes with an environmental focus, for example, on climate change. In those cases where a technology programme does not have any environmental aim among its objectives, all environmental aspects of the programme may be completely unnoticed. Thus the final report of the Pigments programme (1998-2002) does not include any information on environmental aspects (Tekes 2002). (Article V)

Environmental policy integration at the output level, i.e. in the research contracts actually signed is rather weak. Tekes's (2004a) application forms for project support do not have any section specifying the environmental impacts, despite these being mentioned as funding criteria in Tekes's strategy. Even though funding applications and decisions are confidential, which makes the evaluation more difficult, verbal comments by authorities and the results of a questionnaire to financed research teams suggest that positive environmental impacts are indeed counted as a plus if mentioned in the application. But no systematic assessment of the overall environmental impacts of the proposed projects, including positive as well as negative aspects, is made. Another aspect impeding integration is the lack of reporting and evaluation requirements for the projects (Kivimaa and Mickwitz 2005). (Article V)

Specific innovations with less environmentally harmful effects would be the ultimate measure of policy integration. Innovations, however, usually result from a variety of driving forces, of which technology policies are only one (Kivimaa and Mickwitz 2004). An innovation may be an outcome of several consecutive technology programmes, since technology development takes time. Nevertheless some assessments of programme outcomes have been made (Kivimaa and Mickwitz 2005).

In the 1990s, the water consumption of Metso paper machines declined from 7.5 m^3 per tonne of paper to 2.5 m^3, partly due to research carried out in the CACTUS programme (Haavanlammi 2001). Developments partly attributed to the Sustainable Paper Programme included methods enabling reductions of energy consumption by 5-15 per cent in mechanical pulping processes (Lähepelto 1998). POM technology – acclaimed one of the most significant pulp and paper innovations of the 1990s – enables improved energy efficiency and reduced water consumption. One phase of the development of POM took place as part of the CACTUS programme, but even though the programme had environmental aims they did not affect the financing of this technology since its environmental side-effects were unanticipated (Kivimaa and Mickwitz 2004).

As a summary of the evaluated policy integration in Finland, one can say that environmental concerns are reasonably well integrated into technology policy strategies. In the technology programmes and the R&D projects actually financed, there was much less integration. For the technology programmes, one could even speak about an intentional disintegration, since the focus has been on separate programmes for environmental technologies, while environmental issues have been integrated into other programmes only when positive impacts are recognised in the application process. (Article V and Kivimaa and Mickwitz 2004)

7. THE USE OF ENVIRONMENTAL POLICY EVALUATIONS

7.1. DIFFERENT TYPES OF USE OF EVALUATIONS

An aspect in the definition of evaluation promoted by Vedung (1997, 3) and also in the definition utilised in this thesis is that it *"is intended to play a role in future, practical action situations"*. But what practical role, or actually roles, could evaluations of environmental policies play? Although justifications for undertaking evaluations of environmental policies are generally based either on learning and development or on accountability (Section 3.1) the potential roles of evaluations are much broader. Undertaking an evaluation almost always implies some anticipated use to be made of it, but this, of course, does not guarantee any use. The use made of an evaluation and its results neither can nor should be controlled by the evaluators, but at the same time it is clear that the evaluation design and practices will always affect its use. (Article I)

Evaluations of environmental policies are often assumed to affect the implementation of these policies, their design, the design of new policies and the debate on environmental issues as well as public policy in general. That is, they are presumed to ultimately induce single as well as double loop learning (Section 3.1, Argyris 1999, Leeuw et al. 2000).

After evaluators and policy scientists realised that single loop learning directly induced by any evaluation was uncommon, or even rare, there were two responses. On the one hand a lot was done to increase this type of learning by developing the evaluation practices, and on the other hand more focus was put on the role of evaluations for conceptual use and indirect double loop learning. (Article I)

Within the evaluation context, the concept "conceptual use" or "enlightenment" has been used to describe the use of evaluations to broaden minds without any requirement for direct action (e.g. Weiss 1998, Shulha and Cousins 1997). When views and perceptions are formed in pluralistic deliberation conceptual use takes place. Surely no conceptual use of any evaluation can occur in isolation; rather, it is combined with a lot of prior knowledge and information from other sources (Shadish et al. 1995, Albæk 1995).

Weiss (1998) refers to "Use for persuasion" for use in order to mobilize support for positions already held as well as legitimizing use by those with power. Politicians as well as bureaucrats

may commission evaluations only to legitimize policies in place, their implementation practices and their expenses. As discussed, the demand for environmental policy evaluations is not only based on the wish to obtain support for new policies, political support is also needed to maintain present environmental policies in force. Democracies are characterised by conflicting views, and thus evaluations – the processes as well as the results – will be used to support and refute arguments by different individuals and groups in the political process (Albæk 1995, Sabatier 1988, Weiss 1998).

While the use of the findings and recommendations of an evaluation surely is the first aspect one might consider, other aspects of an evaluation may also be used. Especially in participatory evaluations the process itself may be used for dialogue and learning. Concepts, ideas and generalizations developed in an evaluation may be transferred and used in other contexts. The very fact that a policy is under evaluation may be used to legitimize the policy style. (Article I)

The evaluation of environmental policy instruments (reported in Articles II and IV) has been used in both public discussions in the media and expert discussions in the Finnish administration about such issues as selection of policy instruments, regulatory control and innovations, and the role of transparency for policy efficiency. It has thus been one element forming the general knowledge base on which people form their views and opinions on these issues, but still just one element and probably a small one. The evaluation, however, also formed a base against which the new Environmental Protection Act was evaluated (Section 5.5). The evaluation concepts used were seen as useful enough by the Ministry of the Environment and other key interest groups to cause them to commission a follow-up based on the same concepts. The results also constituted a base-line against which the rather limited experiences obtainable during the two first years of the new act could be compared. This allowed the subsequent evaluation to also be used for single loop learning. It was on one hand used to increase the resources of regional permit authorities, but also to initiate further concrete examinations in which activities should be regulated by case-by-case permits, i.e. to review the list contained in the first paragraph of the Environmental Protection Decree (169/2000).

7.2. PROMOTING THE USE OF EVALUATIONS

The frequently observed (e.g. Patton 1997, 7-10) lack of instrumental use, or single loop learning resulted in a wide range of analysis of how evaluators could promote the use of their evaluations. Some evaluation theorists have seen utilisation as such a crucial issue that they have formed what one could call a branch of evaluation based on it. Patton (1997, 20) makes this very explicit stating *"**Utilization-Focused Evaluation** begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done **from beginning to end**, will affect use. …the **focus** in utilization-focused evaluation is on **intended use by intended users**.* [emphases in the original]*."*

That evaluators try to promote the use of their evaluations is necessary or even mandatory. It is however, also problematic (e.g. Wildavsky 1984). There is a clear risk that close relationships with potential users may influence the findings of the evaluation. People tend to use findings in line with their basic belief systems more than conflicting ones (e.g. Sabatier 1988), a heavy emphasis on use may therefore result in more and more evaluations that just confirm already established positions.

Another risk is linked to the interpretation of users, or "intended users". When some groups and individuals are involved in the planning and execution of an evaluation, it usually means that others are not participating. Even the groups involved can seldom be equal. Utilization-centred evaluation tends to strengthen the role of the group(s) commissioning the evaluation. In addition, even in cases when participation is open, not all potential users have the same opportunities to get involved. Different groups have different resources, such as members, money, personnel or intellectual capabilities, these resources have implications for the nature and the extent of their participation. Finally, not all interests are organised at all and it may be questionable which groups should represent a certain interest. Evaluating policies dealing with those environmental problems that affect geographically remote regions pose additional challenges when participation should be ensured. (Article I)

Realising the potential problems of involving stakeholders or potential users in an evaluation does not mean that they should not be involved. Evaluation approaches without any involvement are often far more problematic. While involving only those commissioning an evaluation, which can seldom be avoided, may result in biases that can at least partly be counterbalanced through broader participation. The evaluation of the new Environmental

Protection Act was conducted in close interaction with the Working Group for Assessing the Implementation of the Environmental Protection Act. This ensured that not only civil servants from the Ministry of the Environment, but also representatives of the regional authorities, the municipalities, key interest organisations such as the Confederation of Finnish Industries (EK) and the Central Union of Agricultural Producers and Forest Owners (MTK) and the Finnish Association for Nature Conservation (SLL), could comment on methods and data as well as results during the evaluation process. Furthermore, promoting widespread utilisation of evaluations requires great transparency of both evaluation processes and results. As stated by principle E.3 in the Guiding Principles for Evaluators of the American Evaluation Association (2005, or any other issue of the American Journal of Evaluation): *"Evaluators articulate and take into account the diversity of general and public interests and values and thus should: … 3. Allow stakeholders access to and actively disseminate, evaluative information and present evaluation results in understandable forms that respect people and honour promises of confidentiality."*

Albæk (1996, 17) points out three different biases in agency-sponsored evaluations: selection of evaluators; determination of scope, design and method; and modification of results to please vested interests. An example he uses is an evaluation focused by Danish authorities so that it would demonstrate success and legitimize their work. That the Finnish Ministry of the Environment co-financed the evaluation of the policy instruments used in the pulp and paper and chemical industries, on which this thesis is based, may have been partly for purposes of legitimizing. It was well known that water discharges from the paper and pulp industry had diminished and thus the pre-assumption that policies had played a positive role was not far fetched. Although there is nothing to indicate that there was not also a genuine interest to learn and develop the policies, undertaking the evaluation could also be used to create an image of the Ministry as serious and modern and to respond to the criticism for the Ministry's slight interest in evaluations (Harrinvirta et al. 1998).

8. FOUNDATION OF ENVIRONMENTAL POLICY EVALUATION: BETWEEN CONSTRUCTIVISTS AND REALISTS

In evaluation theory, as in social sciences more generally, there has been an intense debate between different fundamental approaches to ontology and epistemology. Even though this thesis is not about the philosophical foundations of environmental policy evaluation, or evaluation in general, a brief discussion of the concepts and practices promoted in relation to this debate will enrich the views of evaluation along these lines. For this purpose a rough representation of the different perspectives is presented.

Since both "constructivists" and "realists", in the context of evaluation, were partly a response to the quasi-experimental evaluation approach promoted in the early days of evaluation, the debate between them is inseparable from this starting point. Based on the view that public policies should be based on a continuous process of tryouts, assessing effectiveness, adjusting and spreading successful approaches, Donald T. Campbell promoted experimental or quasi-experimental evaluation designs (Shadish et al. 1991, 119-170).

The best known advocates for a constructivist evaluation perspective are Guba and Lincoln[8] (e.g. 1989). They state that *"there is no objective truth on which inquiries can converge"* and thus *"one cannot find out how things really are or how they really work"* (Guba and Lincoln 1989, 46). They argue that the consequence for evaluation is not that anything goes, but to focus on *"how one can compare one construction with another to determine which is to be preferred."* (Guba and Lincoln 1989, 47) An evaluation should thus depart from the claims, concerns and issues of stakeholders. These should be confronted with each other as a basis for learning. Consensus should be generated on as many constructions as possible and then the evaluation should mediate negotiations based on information on those issues where there is no consensus (Guba and Lincoln 1989, 50-57, 72-74).

At the end of the 1990s the focus of the debate in evaluation theory shifted towards realistic evaluation. Realistic evaluation emerged as a critic of both experimental evaluation and constructivist evaluation (e.g. Pawson and Tilley 1997, 4-23)[9]. The realists argue that policies

8 Although Guba's and Lincoln's (1989) evaluation theory is generic, the practical examples are mainly from the area of educational evaluations.

9 Pawson and Tillew (1997) present a generic evaluation theory, but their examples are mainly related to the evaluation of crime prevention.

may induce change. When they do so it is because there is some mechanism between the intervention and the subject it affects (Pawson and Tilley 1997, 65-69, 71-75). Interventions, however, do not always result in the same change, because of the influence of the context (Pawson and Tilley 1997, 76-77). The focus of realistic evaluation has been synthesised in the question *"what works for whom in what circumstances?"* (Pawson and Tilley 1997, Pawson 2003, 474). Realistic evaluation should depart from making programme-specific theories framed as *"propositions about how mechanisms are fired in context to produce outcomes"* (Pawson and Tilley 1997, 84-85). Based on these propositions hypotheses should be made and tested through multi-method data collection and analysis (Pawson and Tilley 1997, 85). Finally the theory of the interplay between mechanisms, context and outcomes should be revised and accumulated in families of theories (Pawson and Tilley 1997, 86).

Almost all the concepts suggested for environmental policy evaluation could be utilised within the Guba and Lincoln constructivist framework. Emphasising side-effects and endorsing the expansion of criteria are useful ways of promoting many, rather than few views on the merits, worth and value of the policy evaluated. While focusing on evaluating intended effects and effectiveness as goal achievement would imply being constrained to only the problem definitions by those in power when the policy was adopted. Stressing the role of several intervention theories as reconstructions of different assumptions of how policies are supposed to function is based on the very idea of alternative constructions of the same policy (on programme theory and constructivism see Dahler-Larsen 2001).

Some of the suggested concepts and practices, however, are not entirely in line with Guba's and Lincoln's constructivist approach. For example, triangulation implies moving forward from the discussion of whether to use qualitative or quantitative data and methods to that of utilising both in parallel, complementarily, and emphasising the interaction between the methods. This requires a different point of departure than Guba's and Lincoln's (1989, 259) starting point to *"use primarily, **although not exclusively**, qualitative methods* [emphasis in the original]*"*. Another differentiating aspect is the view on research versus negotiation in evaluation. While Guba and Lincoln (1989, 255) write *"it is not more research that is needed but more negotiation"* it may actually be that the main contribution an evaluation can provide is the opportunity, or sometimes need, to construct new views based on new empirical data and analyses. The role of evaluation as negotiation can be linked to whether one perceives evaluation *as* the policy process or as *part of* the policy process.

The programme-specific theories of the realistic evaluation are part of the same programme theory-based evaluation approaches that have also inspired the intervention theory concept advocated. There are thus naturally many similarities. The main difference is related to the distinction between *the* programme theory (e.g. Pawson 2003, 474) and strong emphasis on multiple intervention theories proposed (Section 4.4). Also, Pawson and Tilley (1997) use some examples of more than a single theory, but the emphasis is largely on the theory. The programme theories in realistic evaluations can be derived based on views by stakeholders and based on documents, but results accumulated through earlier research and evaluations are also used. Thus the theories reflect propositions about the mechanisms in the specific context based on the synthesis of knowledge prior to the empirical evaluation. Whereas the intervention theories proposed here may well include theories based on assumptions having no support in any research as long as they reflect views held by somebody.

Guba and Lincoln (1989, 96-98) forcefully reject the notion of causality. They *"reject the traditional concept of causality* [which leads them] *to replace it with a different human construction, that of 'mutual simultaneous shaping'"* (Guba and Lincoln 1989, 97) The position of Pawson and Tilley (1997, 58) on causality is also very clear: *"causal outcomes follow from mechanisms acting in contexts"*, which is labelled the *"base upon which all realist explanation builds"*. Although there may be assumptions about causality, even if no causality existed, and intervention theories could be used to reflect this, using intervention theories to plan and conduct empirical evaluations makes more sense in connection to the realistic position on causality.

Finally, a comment on the context specific role of policies, evaluation and policy learning: Guba and Lincoln are very clear: *"generalizations are* ***not possible****"* [emphasis in the original] (1989, 36) and more specifically *"Phenomena can be understood only within the context in which they are studied; findings from one context cannot be generalized to another; neither problems nor their solutions can be generalized from one setting to another."* (1989, 45) Pawson and Tilley (1997, 86), on the other hand, clearly argue for generalization, not unconditionally, but rather as *"specifying those 'regularities' or 'outcome patterns' which the present state of our understanding of 'mechanisms' and 'context' is able to sustain."*

The entire idea of this thesis rests on the view that something may be generalised from one setting to another. Otherwise it would make no sense at all to study evaluation concepts. Since, without any generalisation, all evaluation concepts would have to be constructed in the specific context of every evaluation. This thesis, however, has been based more specifically on

generalisation, because it started out from the assumption that some shared characteristics of environmental problems (Section 2.3) have an impact on the concepts and practice of environmental policy evaluation. Surely these characteristics are not common to all environmental problems and every problem has it own context specific features. As discussed in the case of climate change and eutrophication caused by industrial waste water discharges, both can be described through these characteristics, while their realisation is different in the two cases. However, since characteristics such as complexity and long time frames tend to make the problems, adopting suitable policies to address the problems, and the evaluation of the policies more difficult, it is better to work on the assumption that these shared characteristics are present and then revise the assumption based on the context-specific features rather than the other way around.

Generalisation of concepts and approaches is of course one thing and generalisation of results is another. One could therefore support the concepts and practices proposed as starting points for the evaluation of environmental policies in many different contexts and continue to hold the position that the findings from these evaluations have no relevance in any other situation. While it is clear that no policy panacea exists, the perceptions presented above on the use of evaluations (Section 7) are based on the view that experiences from the evaluation of one policy can also be an input in the learning process regarding other policies. This does not imply that context-specific characteristics could be overlooked. Learning based on findings from other contexts should always be cautious.

9. CONCLUSIONS

This thesis – the summary as well as the articles – was motivated by the increased demands for evaluations of environmental policies. This summary started with two questions: *"Should environmental policies be evaluated?"* and if so *"How could they be evaluated?"*. The argument evinced with respect to the first question is that the demand for evaluations should indeed be met, since evaluations can be a source for learning and development, as well as an important way to promote accountability. The second question was addressed by examining potential concepts for evaluations of environmental policies. Side-effect evaluation, multi-criteria evaluation, intervention theories and triangulation have been conceptually discussed as well as tested in practice and found to provide useful means for the evaluation of environmental policies.

While environmental policies do need to be evaluated, the focus should not be restricted to merely assessing effectiveness. A broader perspective is required. The merits of a policy can never depend solely on its ability to meet its goals. Useful evaluations of environmental policies are not about providing simple verdicts, but about providing many kinds of inputs for the political processes through which environmental policies are formulated, reformulated and implemented. A multi-criteria approach is more likely to offer such inputs than a single criterion evaluation, such as an effectiveness evaluation.

Case-by-case permits were found to have many positive impacts while regulating pollution from large-scale point sources, such as pulp and paper mills. The implementation of the systems, however, took a long time, especially when new acts came into force, but also when new content requirements were introduced. The permits are thus most suitable for environmental problems on which there is longer time to act. The flexibility provided by case-by-case decision making has most advantages when the differences between local environmental conditions, technologies, costs, etc. are large, as in Finnish pulp and paper production. That the value of flexible permitting depends both on the context and the evaluation criteria was established when contrasting these experiences with those from applying new environmental permits to sites such as petrol stations and stone crushing plants.

The learning and development that can be promoted through evaluations, is not only related to improving already existing policy instruments; it is also linked to selecting new instruments. There has since the 1990s been widespread enthusiasm about using economic and information-based environmental policy instruments rather than regulation. This preference

is often based on not taking all aspects into account, aspects that in the case of Finnish permit were revealed by multi-criteria. Furthermore, the mechanisms through which traditional regulations work are not always well known; an ignorance that tend to underestimate their impacts.

The evaluation of the Finnish permit system stressed the role of the mills' ability to anticipate gradually expanding and tightening environmental requirements. Companies do not act on the basis of today's requirements alone, especially not when making far-reaching decisions concerning, among other things, investments or research and development. An important feature of all types of environmental policy instruments is therefore predictability. While predictability at a general policy level may have some effect, it is often the ability to foresee the policies as implemented that promotes early or innovative action. Examples may equally well be expectations about specific limit values as about the existence or non-existence of exemptions from environmental taxes. For future requirements to have effects knowledge is not enough; declared policy intentions need to be credible. Predictability should be a more important criterion than hitherto when selecting policy instruments and planning their implementation.

Addressing environmental problems is often challenging due to long time frames, geographical distribution of causes and effects, complexities, as well as features related to our knowledge of environmental problems. These aspects not only pose challenges to policies but also on their evaluation. If they are not taken seriously into account there is a risk of biased evaluations. The bias would systematically tend to be in the direction of little or no effects of policies, and the ensuing policy implications could easily be either "laissez-faire" or deregulation.

Policies, including environmental policies, may actually have intended effects. But due among other things to complexities and limited knowledge they often also have unanticipated effects. This aspect has long been noted in the case of environmental side-effects of other policies and is thus the key justification for the demands of environmental policy integration. But environmental policies also tend to produce unanticipated effects and one of the potentially most important contributions of evaluations could be through shedding new light on these.

Politics is about promoting conflicting objectives, but it is also about acting under different beliefs and assumptions. Formulating and implementing policies is inseparable from politics and the different assumptions and beliefs are therefore part of the policies. Intervention theories, constructed on the basis of different assumptions about relevant actors, inputs,

outputs and outcomes as well as about how these are linked, can be useful tools to reveal the differences in assumptions and beliefs. But intervention theories can also be used to relate these assumptions and beliefs to empirical data and analysis.

Although there are many advantages related to the evaluation concepts – side-effect evaluation, multi-criteria evaluation, intervention theories and the triangulation based approach – proposed in this thesis there are also drawbacks. The greatest challenge is that the approaches tend to make evaluations broader. This is almost self-evident when criteria, materials and methods are increased and the focus is extended from main effects to cover side-effects. Broader evaluations require more resources if the depth of the analyses is to be kept on the same level. In addition, a wider set of skills is often needed. Since few people can conduct all types of analyses, the obvious response is larger evaluation teams, with a variety of backgrounds and skills. To combine a wide evaluation approach with a large and diverse evaluation team is a difficult but not impossible task. It is, however, especially important that communication challenges are recognised and that sufficient time and opportunities for interaction are provided throughout the evaluation process.

The political background of evaluating environmental policies has been discussed in this thesis. The use of evaluations has been discussed in relation to political processes and the general demands for evaluations have been examined in the light of "new public management". In Finland the transfer of duties to regional agencies and municipalities and the new budgeting approaches are examples of such reforms that have influenced the requests for environmental policy evaluations. There is, however, much additional research to be done on the politics of evaluation. An analysis of the role of the rise of right wing parties in many countries, e.g. Denmark, when cost-effectiveness evaluation and "more environment for the money" become important evaluation drivers is one avenue for further research.

This thesis has provided concepts and demonstrated that they can be used in practice. Concepts and practices, however, are not the only aspects of an advanced evaluation culture. There are also institutional issues. Should the evaluation requirement of EU environmental policies be operationalized by the EU Commission commissioning evaluations? Or by the Council, the Parliament, the Member States or perhaps by the European Environmental Agency? Should there be some institution with specific capacities to conduct the evaluations, as in some countries and sectors and what are the pros and cons of different structures? Similar questions are also important at the national, regional and local level. These issues, however, are beyond the scope of this research.

In this thesis four evaluation concepts were discussed: side-effect evaluation, multi-criteria evaluation, intervention theories, and triangulation. While they have mainly been discussed separately in this thesis, they are linked and one could even say that together they form an evaluation model. In order to provide a broad picture of environmental policy the side-effect evaluation perspective is a particularly useful starting point, emphasising that a key task is to draw a more complete picture of all the effects. In order to make statements on the merit, worth and value of the policy, the effects as such are not enough; some criteria are also required. Fruitful deliberation is better advanced if several criteria are used instead of just one. Intervention theories constitute a tool that can be used in order to plan the evaluation in relation to the anticipated effects. It is thus appropriate to retain the side-effect evaluation as the organising principle and to consider intervention theories as a concept of use with respect to a subset of the effects, that is, the anticipated effects. Finally, some empirical work has to be undertaken in order to explore effects, utilise criteria and examine intervention theories. Here the role of triangulation is emphasised. By utilising several data sources, methods, theories and complementary perspectives of several evaluators one is more likely to discover unanticipated effects as well as to be able to assess the always complicated causal relationships.

ACKNOWLEDGEMENTS

This thesis, the summary and all the articles, was written as part of my work at the Finnish Environmental Institute, until the Institute was reorganised at the Environmental Policy Instruments Division, and, since 2002, in the Research Programme for Environmental Policy. The conditions for learning about environmental policies and doing research have been excellent and it has been a privilege to be part of such a working community.

All the research on which this thesis is built was conducted on projects led by Professor Mikael Hildén. He not only made this thesis possible, but also provided much input for its content. He was the first with whom many of the ideas were discussed, he was the co-author of Article IV, he provided detailed comments on the draft summary – from Africa – and he gave me advice on some final additions as late as last week. For all this I will always be grateful, but my debt is even greater; had we not met, I would most likely never have started to do research in environmental policy.

Professor Yrjö Haila suggested that I should do my doctoral dissertation at the University of Tampere. Since I started as a doctoral candidate in the autumn 2004 he, *Dosentti* Pekka Jokinen as well as many participants in the "umbrella" seminar have provided a lot of useful comments and much encouragement. Professor James Meadowcroft and Professor Markku Temmes were the pre-examiners of the thesis, the insights contained in their reports and the recommendations for both improvements and further projects were most helpful.

Professor Evert Vedung introduced me to much of the evaluation literature I know, he has given detailed comments on four of the articles as well as the summary and we have had numerous discussions that have enhanced my understanding of evaluation, public policy and life. Co-operation with Dr. Mikael Román, Professor Kenneth Oye, Dr. James Foster, Dr. Audun Ruud and Professor William Lafferty have increased my understanding of regulation and environmental policies.

Paula Kivimaa was the co-author of Article V, this is only one out of four papers we have jointly written in recent years. At least from my perspective our co-operation has been easier, more intense, more efficient, but at the same time more rewarding than any collaborative research I have experienced before I truly hope this is just the beginning of our teamwork. My writing style as well as my understanding of the publishing processes has been greatly improved through the publishing of three articles with Professor Matti Melanen. Virginia Mattila proofread several of the articles as well as the summary, she has made my English much more understandable for others as well as for me.

My exploration of environmental policy evaluation would have been much duller and less revealing if I had not had the opportunity to make it together with many colleagues. The chance to talk and write with Jukka Similä has been a key to what I know about regulation and Finnish environmental law in particular. Co-operation with Petrus Kautto advanced my understanding of the interaction between companies and policies as well as music, especially the "Kärpät song". Kimmo Ollikka, Elise Sahavirta, Professor Stefan Sjöblom, Ulla Rosenström, Mikko Attila, Alec Estlander, Jouko Petäjä, Heli Hyvättinen, Marianne Lindström, Elina Karhu and Heidi Vuoristo have also all provided very valuable help on various evaluation projects. Colleagues not directly involved in joint projects have also offered much help especially Dr. Helena Valve, Pauliina Jalonen, Panu Kontio and Jouko Tuomainen.

Another important source for knowledge of concepts as well as practices of evaluation has been the Finnish Evaluation Society (FES), and above all its board. Since 2001, when I became a member of FES board I have learnt something new at almost every meeting I have attended. For example, Section 8 is partly a response to a comment by Professor Pertti Ahonen that all evaluators should have a declared position between constructivists and realists. Although all board members have offered me something I would in particular like to thank; Johanna Nurmi, Veli-Pekka Sinervuo, Dr. Riitta Seppänen-Järvelä, *Dosentti* Liisa Horelli and *Dosentti* Petri Virtanen.

General Director Lea Kauppi has transformed the Finnish Environmental Institute into a serious multidisciplinary research institute, where social sciences are also appreciated and research like in this thesis is possible. Research manager Eeva Furman was the one who finally pushed me to undertake this dissertation, today I am very happy about that.

I am very grateful that The Finnish Society of Sciences and Letters agreed to publish my thesis in the series Commentationes Scientiarum Socialium. The editor-in-chief of the series Per Schybergson not only accepted the manuscript but also made several valuable editorial comments. Tiina Kaarela took on the task of doing the layout of this publication with little time and at a time – just before Christmas – when one should not be under pressure. Four of the articles were written on projects that have obtained funding from the Finnish Environmental Cluster Research Programme and one from the Research Programme for Advanced Technology Policy (ProACT).

Finalising this thesis has also demanded some characteristics that are not my personal strengths; notably patience but also writing formal applications and filling in forms. At times this has affected my temper and I have not always been the constructive colleague I would

wish to be. I apologise, and especially to Eeva. Those, however, who have suffered the most both from my occasional irritability and more frequent mental absence are my family: Anna-Karin, Valter and Theo, thank you for coping with me.

Helsinki, November 30th 2005

Per Mickwitz

REFERENCES

Ahonen Pertti, Petri Virtanen and Petri Uusikylä 2002. "Evaluation in Finland", in Furubo Jan-Eric, Ray C. Rist and Rolf Sandahl (eds.) *International Atlas of Evaluation*. Comparative Policy Analysis Series. pp. 47-61. New Brunswick: Transaction Publishers.

Albæk Erik 1995. "Between Knowledge and Power: Utilization of Social Science in Public Policy Making". *Policy Sciences* 28(1): 79-100.

Albæk Erik 1996. "Why All This Evaluation? Theoretical Notes and Empirical Observations on the Functions and Growth of Evaluation, with Denmark as an Illustrative Case". *The Canadian Journal of Program Evaluation* 11(2): 1-34.

American Evaluation Association 2005. "Guiding Principles for Evaluators". *American Journal of Evaluation* 26(3): 297-8.

Andersen Mikael Skou and Duncan Liefferink 1997. "Introduction", in Andersen Mikael Skou and Dunkan Liefferink (eds.) *European environmental policy: The pioneers*. pp.1-39. Manchester: Manchester University Press.

Argyris Chris 1999. *On Organizational Learning*. Second Edition. Oxford: Blackwell Publishing.

Banuri Tariq, Karl-Göran Mäler, Michael Grubb, Harold K. Jacobson and Farhana Yamin (1996) "Equity and Social Considerations", in Bruce James P., Hoesung Lee and Erik F. Haites (eds.) *Climate Change 1995: Economic and Social Dimensions of Climate Change*. pp. 80-124. Cambridge: Cambridge University Press.

Bartlett Robert 1994. "Evaluating Environmental Policy Success and Failure", in Vig Norman J. and Michael E. Kraft (eds.) *Environmental Policy in the 1990s: Towards a New Agenda*. 2nd edition. pp. 167-197. Washington D.C: Congressional Quarterly Press.

Bouckaert Geert, Derry Ormong and Guy Peters 2000. *A Potential Governance Agenda for Finland*. Research Report 8/2000. Helsinki: Ministry of Finance.

Braden John B. and Chang-Gil Kim (1998) "Economic Approaches to Evaluating Environmental Programs", in Knaap Gerrit J. and Tschangho J. Kim (eds.) *Environmental Program Evaluation: A Primer*. pp. 203-37. Urbana: University of Illinois Press.

Browne Angela and Aaron Wildavsky 1984. "What Should Evaluation Mean to Implementation?", in Jeffrey L. Pressman and Aaron Wildavsky *Implementation: How Great Expectations in Washington are Dashed in Oakland*. Third expanded edition. pp. 181-205. Berkeley, CA: University Of California Press.

Cederlöf Jan-Magnus 2001. *Ecological Modernisation and Market-Based Policy Instruments – The Use of New Instruments in Environmental Policy in Finland and Sweden*. Helsinki: Publications of the Swedish School of Economics and Business Administration.

Chelimsky Eleanor 1995. "Where we stand today in the practice of evaluation: Some reflections". *Knowledge & Policy* 8(3): 8-19.

Chen Huey-Tsyh 1990. *Theory-Driven Evaluations*. Newbury Park, CA: Sage Publications.

Commission for Sustainable Consumption and Production 2005. *Getting more from less – Proposition for a National Programme on Sustainable Consumption and Production* [Vähemmästä enemmän ja paremmin: Kestävän kulutuksen ja tuotannon toimikunnan (KULTU) ehdotus kansalliseksi ohjelmaksi]. Helsinki: Ministry of Environment and Ministry of Trade and Industry. (English version available at: http://www.ymparisto.fi/download.asp?contentid=40471&lan=en accessed 18.11.2005)

Dahler-Larsen Peter 2001. "From Programme Theory to Constructivism: On Tragic, magic and Competing Programmes". *Evaluation* 7(3): 331-349.

Davies Clarence and Jan Mazurek 1998. *Pollution Control in the United States – Evaluating the System*. Washington, DC: Resources for the Future.

Economist 2003. "CO_2 Emissions". *The Economist* May 10, 367(8323): 94.

Environmental Assessment Institute 2005. Strategy Plan 2005 – 2008, Draft of 2 March 2005. (available at http://www.imv.dk/Files/Filer/Strategiplan_mv/SP_DRAFT_02.03.05.pdf accessed 3.3.2005)

European Commission 2000. Global assessment Europe's environment: what directions for the future?. Luxembourg: Office for Official Publications of the European Communities. (available at http://europa.eu.int/comm/environment/newprg/99543_en.pdf accessed 3.3.2005)

European Commission 1999. Global Assessment of the Fifth Environmental Action Programme: Views of sectoral groups/NGOs. (available at http://europa.eu.int/comm/environment/newprg/ngosects.pdf accessed 3.3.2005)

European Environment Agency 2001. *Reporting on Environmental Measures: Are We Being Effective?*. Environmental Issue Report No 25. Copenhagen: European Environment Agency.

European Parliament and the Council of the European Union (2002) Decision No 1600/2002/EC of the European Parliament and of the Council of 22 July 2002 Laying Down the Sixth Community Environment Action Programme. *Official Journal of the European Communities* L242(45): 1-15.

Finnish Government Working Group for Planning and Steering of the Development of Law-Drafting 2005. *Tehokkaampaa, suunnitelmallisempaa ja hallitumpaa lainvalmistelua* [A More Efficient, Better Planned and More Controlled Law-Drafting]. Prime Minister's Office Publications 13/2005. Helsinki: Prime Minister's Office.

Furubo Jan-Eric and Rolf Sandahl 2002. "Introduction: A Diffusion Perspective on Global Developments in Evaluation", in Furubo Jan-Eric, Ray C. Rist and Rolf Sandahl (eds.) *International Atlas of Evaluation*. Comparative Policy Analysis Series. pp. 1-23. New Brunswick: Transaction Publishers.

Gore Al 1996. *The Best Kept Secrets in Government: How the Clinton Administration Is Reinventing the Way Washington Works*. National Performance Review. Washington, D.C.: Government Printing Office.

Guba Egon G. and Yvonna S. Lincoln 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage Publications.

Haavanlammi Tuija 2001. Shower Water System Under the Cactus Program, Comments by Metso Oyj. (Available at: http://www.tekes.fi/julkaisut/cactus_final/results/companies/index.html accessed 29.9.2004)

Haila Yrjö 1998. "Environmental problems, ecological scales and social deliberation", in Glasbergen Pieter (ed.) *Co-operative Environmental Governance: Public-Private Agreements as a Policy Strategy*. pp. 65-87 Dordrecht: Kluwer Academic Publishers.

Harrinvirta Markku, Petri Uusikylä and Petri Virtanen 1998. *Arvioinnin tila valtionhallinnossa* [The State of Evaluation in the Governmental Administration]. Research Reports 7/98. Helsinki: The Ministry of Finance.

Hildebrand Philipp 2002. "The European Community's Environmental Policy, 1957 to 1992: From Incidental Measures to an International Regime?", in Andrew Jordan (ed.) *A Guide to EU Environmental Policy: Actors, Institutions and Processes*. pp. 13-36. London: Earthscan.

Hildén Mikael, Mikko Attila, Marjukka Hiltunen, Niko Karvosenoja and Sanna Syri 2001. *Kansallisen ilmastostrategian ympäristövaikutusten arviointi* [Environmental Impact Assessment of the National Climate Strategy]. Suomen ympäristö 482. Helsinki: Finnish Environment Institute.

Hildén Mikael, Jukka Lepola, Per Mickwitz, Aard Mulders, Marika Palosaari, Jukka Similä, Stefan Sjöblom and Evert Vedung 2002. *Evaluation of Environmental Policy Instruments – A Case Study of the Finnish Paper & Pulp and Chemical Industries*. Monographs of the Boreal Environment Research 21. Helsinki: Finnish Environment Institute.

Hildén Mikael, Petrus Kautto, Per Mickwitz and Jukka Similä 2003. "Ympäristönsuojelulain kaksi ensimmäistä vuotta – arvioinnin yhteenveto" [The Two First Years of the Environmental Protection Act – A Summary of the Evaluation]. *Ympäristöjuridiikka* 24(1): 181-199.

Hill Michael and Peter Hupe 2002. *Implementing Public Policy: Governance in Theory and in Practice*. London: SAGE Publications.

Hoogerwerf Andries 1990. "Reconstructing policy theory". *Evaluation and Program Planning* 13 (3): 285-291.

Intergovernmental Panel on Climate Change (IPCC) 2001. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson [eds]). Cambridge: Cambridge University Press. (Also available at: http://www.grida.no/climate/ipcc_tar/wg1/index.htm accessed 14.5.2003).

Jacob Klaus and Axel Volkery 2003. *Instruments for Policy Integration: Intermediate Report on the RIW project POINT*. Environmental Policy Research Centre (FFU), FFU-report 06-2003. Berlin: Environmental Policy Research Center.

Jung Chulho, Kerry Krutilla and Roy Boyd 1996. "Incentives for Advanced Pollution Abatement Technology at the Industry Level: An Evaluation of Policy Alternatives". *Journal of Environmental Economics and Management* 30 (1): 95-111.

Joas Marko 1997. "Finland: from local to global politics", in Andersen Mikael Skou and Duncan Liefferink (eds.) *European environmental policy: The pioneers*. pp. 119-160. Manchester: Manchester University Press.

Jänicke Martin and Helmut Weidner 1995. "Successful Environmental Policy: An Introduction", in Jänicke Martin and Helmut Weidner (eds.) *Successful Environmental Policy – A Critical Evaluation of 24 Cases*. edition sigma. pp. 10-26. Berlin: Rainer Bohn Verlag.

Kemp René 1997. *Environmental Policy and Technical Change – A Comparison of the Technological Impact of Policy Instruments*. Cheltenham: Edward Elgar.

Kivimaa Paula and Per Mickwitz 2004. "Driving Forces for Environmentally Sounder Innovations: The Case of Finnish Pulp and Paper Industry", In: Klaus Jacob, Manfred Binder and Anna Wieczorek (eds.) *Governance for Industrial Transformation*. Proceedings of the 2003 Berlin Conference on the Human Dimensions of Global Environmental Change. pp. 356-372. Berlin: Environmental Policy Research Centre.

Kivimaa Paula and Per Mickwitz 2005. Can We Get Greener Technologies Through Greener Policies? Environmental policy integration in Finnish technology policies. Unpublished manuscript.

Knaap Gerrit J. and Tschangho J. Kim 1998a. "Introduction: Environmental Program Evaluation: Framing the Subject, Identifying Issues", in Knaap Gerrit J. and Tschangho J. Kim (eds.) *Environmental Program Evaluation: A Primer*. pp. 1-20. Urbana: University of Illinois Press.

Knaap Gerrit J. and Tschangho J. Kim 1998b. "Conclusion: Environmental Program Evaluation: Promise and Prospects", in Knaap Gerrit J. and Tschangho J. Kim (eds.) *Environmental Program Evaluation: A Primer*. pp. 347-60. Urbana: University of Illinois Press.

Kronsell Annica 1997. "Sweden: setting a good example", in Andersen Mikael Skou and Duncan Liefferink (eds.) *European environmental policy: The pioneers*. pp. 40-80. Manchester: Manchester University Press.

Laakkonen Simo, Sari Laurila and Marjatta Rahikainen (eds.) 1999. *Harmaat aallot – ympäristönsuojelun tulo Suomeen* [Grey Waves – The introduction of environmental protection in Finland]. Helsinki: Societas Historica Finlandiae.

Lafferty William M. and Eivind Hovden (2003) "Environmental Policy Integration: Towards an Analytical Framework". *Environmental Politics* 12(3): 1-22.

Lafferty William M. and James Meadowcroft 1996. "Democracy and the environment: congruence and conflict – preliminary reflections", in Lafferty William M. and James Meadowcroft (eds.) *Democracy and the Environment: Problems and Prospects*. pp. 1-17. Cheltenham: Edward Elgar.

Leeuw Frans, Ray C. Rist, and Richard Sonnichsen (eds.) 2000. *Can Government Learn? – Comparative Perspective on Evaluation & Organizational Learning*. New Brunswick: Transaction Publishers.

Leeuw Frans 2003. "Reconstructing Program Theories: Methods Available and Problems to be Solved". *The American Journal of Evaluation* 24 (1): 5-20.

Liberatore Angela 1997. "The integration of sustainable development objectives into EU policy-making: Barriers and prospects", in: Baker Susan, Maria Kousis, Dick Richardson and Stephen Young (eds.) *The politics of sustainable development*. pp. 107-126. New York: Routledge.

Liljelund Lars 2004. "Foreword by the Management Board Chairman", in *EEA strategy 2004–2008*. p. iii. Copenhagen: European Environment Agency.

Lundqvist Lennart J. 1996. "Environmental Politics in the Nordic Countries: Policy, Organisation, and Capacity", in Peter Munk Christiansen (ed.) *Governing the Environment: Politics, Policy, and Organization in the Nordic Countries*. Nord 1996:5. pp. 13-27. Copenhagen: Nordic Council of Ministers.

Lähepelto Jaakko 1998. *Sustainable Paper Research Programme Final Report 1993 –1998*. Helsinki: National Technology Agency Tekes.

Mickwitz Per 1998. *Positive Measures: Panacea or Placebo in International Environmental Agreements*. Nord 1998:11. Copenhagen: Nordic Council of Ministers.

Mickwitz Per 2000. Environmental Policy Instruments and Innovations. Unpublished paper presented at the International Society for Ecological Economics Conference, 5 - 8 July 2000 in Canberra.

Mickwitz Per, Kimmo Ollikka, Stefan Sjöblom and Charlotta von Troil 2003. "Ympäristönsuojelulain mukaiset päätökset kahden ensimmäisen vuoden aikana" [The Decisions According to the Environmental Protection Act During the Two First Years]. *Ympäristöjuridiikka* 24(1): 27-47.

Mickwitz Per, Heli Hyvättinen and Paula Kivimaa 2005. The role of policy instruments for the innovation and diffusion of environmentally friendlier technologies. Unpublished manuscript.

Ministry of the Environment 1989. Decision taken by the Ministry of the Environment of Finland on 22 June 1989 on the reduction of discharges of chlorinated organic compounds from the pulp and paper industry. p. 3.

Ministry of the Environment 1998. *Futures for FEI: International Evaluation of the Finnish Environment Institute.* The Finnish Environment 269. Helsinki: Ministry of the Environment.

Ministry of the Environment 2002. *Ekotehokas yhteiskunta ja hyvinvointia edistävä elinympäristö – Ympäristöministeriön strategia* [An Eco-efficient Society and Welfare Promoting Environment – The Strategy of the Ministry of the Environment]. Ympäristöministeriön moniste 101. Helsinki: Ministry of the Environment.

Ministry of the Environment 2004. *Ympäristöministeriön hallinnonalan keskeiset tavoitteet ja tehtävät vuosina 2006 – 2009* [Key objectives and tasks of the environmental administration during the years 2006 – 2009]. Ympäristöministeriön moniste 143. Helsinki: Ministry of the Environment.

Nagarajan Nigel and Marc Vanheukelen 1997. *Evaluating EU Expenditure Programmes: A Guide – Ex Post and Intermediate Evaluation*. XIX/02-Budgetary overview and evaluation, Directorate-General XIX. (Available at: http://europa.eu.int/comm/budget/evaluation/en/guide/guide00-toc.htm. accessed 9.8.2000).

OECD 2002. *Improving Policy Coherence and Integration for SD: A Checklist*. Paris: Organisation for Economic Cooperation and Development.

Ojala Olli 1997. "Suomen ympäristösuojelun kehittyminen" [The development of environmental protection in Finland]. *Ympäristö ja Terveys* 28(3-4): 72-85.

Patton Michael Q. 1997. *Utilization-Focused Evaluation: The New Century Text*. Edition 3. London: SAGE Publications.

Pawson Ray 2003. "Nothing as Practical as a Good Theory". *Evaluation* 9(4): 471-490.

Pawson Ray and Nick Tilley 1997. *Realistic Evaluation*. London: SAGE Publications.
Peters Guy 1998. "Managing Horizontal Government: The Politics of Coordination". *Public Administration – an international quarterly* 76(2): 295-311.
Peters Guy 2001. *The Politics of Bureaucracy*. Fifth edition. London: Routledge.
Pollitt Christopher and Geert Bouckhaert 2000. *Public Management Reform: A Comparative Analysis*. Oxford: Oxford University Press.
Pollitt Christopher and Geert Bouckhaert 2004. *Public Management Reform: A Comparative Analysis – Second Edition*. Oxford: Oxford University Press.
Popper Karl 2003. *The Open Society and Its Enemies, Volume Two: Hegel and Marx*. London: Routledge Classics.
Portney Paul R. and John P. Weyant (eds.) 1999. *Discounting and Intergenerational Equity*. Washington, DC: Resources for the Future.
Pressman Jeffrey L. and Aaron Wildavsky 1984. *Implementation: How Great Expectations in Washington are Dashed in Oakland*. Third expanded edition. Berkeley, CA: University of California Press.
Rohdes Rod 2000. "Governance and Public Administration", in Pierre Jon (ed.) *Debating Governance: Authority, Steering, and Democracy*. pp. 54-90. Oxford: Oxford University Press.
Rogers Patricia J., Anthony Petrosino, Tracy A. Huebner and Timothy A. Hacsi 2000. "Program Theory Evaluation: Practice, Promise, and Problems", in P. J. Rogers, T. A. Hacsi, A. Petrosino and T. A. Huebner (eds.) *Program Theory in Evaluation: Challenges and Opportunities*. New Directions for Evaluation 87. pp. 5-13. San Francisco, CA: Jossey-Bass.
Rossi Peter H., Howard E. Freeman and Mark W. Lipsey 1999. *Evaluation: A Systematic Approach*. Sixth Edition. Thousand Oaks: SAGE Publications.
Sairinen Rauno 2000. *Regulatory reform of Finnish environmental policy*. Centre for Urban and Regional Studies Publications A27. Espoo: Helsinki University of Technology.
Sabatier Paul A. 1988. "An Advocacy Coalition Framework of Policy Change and the Role of Policy-oriented Learning Therein". *Policy Sciences* 21(1): 129-68.
Science and Technology Policy Council 1990. *Katsaus 1990: Tiede- ja teknologiapolitiikan suuntaviivat 1990-luvulla* [Review 1990: Guidelines for Science and Technology Policy in the 1990s]. Helsinki: Science and Technology Policy Council of Finland.
Science and Technology Policy Council 1993. *Tiedon ja osaamisen Suomi: Kehittämisstrategia* [Towards an Innovative Society: A Development Strategy for Finland]. Helsinki: Science and Technology Policy Council of Finland.
Science and Technology Policy Council 1996. *Suomi: Tiedon ja osaamisen yhteiskunta* [Finland: A Knowledge-based Society]. Helsinki: Science and Technology Policy Council of Finland.
Science and Technology Policy Council 2000. *Review 2000: The Challenge of Knowledge and Know-how*. Helsinki: Science and Technology Policy Council of Finland.
Science and Technology Policy Council 2003. *Knowledge, Innovation and Internationalisation*. Helsinki: Science and Technology Policy Council of Finland.
Scriven Michael 1991. *Evaluation Thesaurus*. Fourth edition. Newbury Park, CA: Sage Publications.
Shadish William R., Thomas D. Cook and Laura C. Leviton 1995. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage Publications.
Shulha Lyn M. and Bradley J. Cousins 1997. "Evaluation Use: Theory, Research, and Practice since 1986". *Evaluation Practice* 18(3): 195-208.
Similä Jukka 2002. "Regulation and its effects on technological innovations". *Environmental Law* 14(2): 143-160.
Similä Jukka and Mikael Hildén 2003. "Yhdennetty ympäristölupa – Ympäristönsuojelulakiuudistuksen vaikutukset" [An Integrated Environmental Permit – The Impacts of the Reform of the Environmental Pollution Legislation]. *Ympäristöjuridiikka* 24(1): 4-26.
Sjöblom Stefan and Charlotta von Troil 2003. "YSL:n paikallinen täytäntöönpano ja PK-yritykset" [Local Implementation of the Environmental Protection Act and SMEs]. *Ympäristöjuridiikka* 24(1): 83-97.
Skodvin Tora 2000. "The Intergovernmental Panel on Climate Change", in Andresen Steinar, Tora Skodvin, Arild Underdal and Jørgen Wettestad (eds.) *Science and politics in international environmental regiemes: between integrity and involvement*. pp. 146-180. Manchester: Manchester University Press.
SOU 2005. *Strategi för hav och kust utan övergödning: Miljövårdsberedningens promemoria*. 2005:1 [A Strategy for Ending Eutrophication of Seas and Coasts] Stockholm: EDITA Norstedt tryckeri AB. (Also available in English at: http://www.sou.gov.se/mvb/pdf/Hav%20och%20kust%20engelsk%20version.pdf accessed 30.11.2005)
State Audit Office 2001. *Itämeren suojelusopimuksen toteuttaminen Suomessa – Helsingin sopimuksen merkitys vesiensuojelun ohjausvälineenä erityisesti maalta peräisin olevan kuormituksen vähentämisessä* [Implementation of the Helsinki Convention on the Protection of the Marine Environment of the Baltic Sea Area in Finland]. Performance audit reports 8/2001. Helsinki: State Audit Office.

State Audit Office 2002. *Ympäristölupien valvonta – erityisesti ilmansuojelun kannalta* [The control of environmental permits particularly with regard to air protection]. Performance audit reports 30/2002. Helsinki: State Audit Office.

Swedish EPA 2000. *The Swedish Environmental Protection Agency: An introduction to the Swedish EPA, its organisation, tasks and working methods*. Stockholm: The Swedish Environmental Protection Agency. (Available at http://www.internat.naturvardsverket.se/documents/swepa/swepadoc/intro/swepa.pdf accessed 3.3.2005)

Swett Brian 2004. "Program Evaluation at the U.S.EPA – The Dual Role of the Evaluation Support Division: Capacity Building & Evaluating Innovations". Presentation for the 2004 Canadian Evaluation Society Conference, May 19. (Available at http://www.epa.gov/evaluate/aea/capacity.pdf accessed 3.3.2005)

Tekes (2002) *Pigmentit paperin raaka-aineena 1998-2001 loppuraportti* [Pigments as raw material for paper 1998-2001 final report]. Technology programme report 5/2002. Helsinki: National Technology Agency Tekes.

Tekes (2004a) Technology strategy and online information on applying for project funding. (Available at: www.tekes.fi accessed 7.4.2004)

Tekes (2004b) *Vuosikertomus 2003* [Annual Report 2003]. Helsinki: National Technology Agency Tekes.

Temmes Markku 1998. "Finland and New Public Management". *International Review of Administrative Sciences* 64(3): 441-456.

Temmes Markku 2000. "Evalueringsverksamhet inom statsförvaltningen i Finland" [Evaluation activities within Finnish government]. *Nordisk Administrativ Tidskrift* 81(2): 148-155.

Underdal Arild 1980. "Integrated marine policy: What? Why? How?". *Marine Policy* 4(3): 159-69.

United Nations 1994. *Agenda 21: earth Summit – The United Nations Program of Action from Rio*. New York: United Nations Department of Public Information.

United Nations 2005. "Millennium Indicator: 'Carbon dioxide emissions (CO2), metric tons of CO2 per capita(UNFCCC-UNDESA/Statistics Division)'". (Available at http://millenniumindicators.un.org/unsd/mi/mi_series_results.asp?rowID=577 accessed 4.3.2005)

Vedung Evert 1997. *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.

Vedung Evert 2001. "Utvecklingsdrag i utvärdering" [Features of the Development of Evaluation]. *Hallinnon Tutkimus* (Administrative Studies) 20(3): 139-143.

Vedung Evert 2004. *Utvärderingsböljans former och drivkrafter* [The wave of evaluation and its driving forces]. FinSoc Working Papers 1/2004. Helsingfors: STAKES.

Vedung Evert and Frans van der Doelen 1998. "The Sermon: Information Programs in the Public Policy Process – Choice, Effects and Evaluation", in Bemelmans-Videc Marie-Louise, Ray C. Rist and Evert Vedung (eds.) *Carrots, Sticks, & Sermons: Policy Instruments & Their Evaluation*. pp. 103-128. New Brunswick: Transaction Publishers.

Weale Albert 1992. *The New Politics of Pollution*. Issues in Environmental Politics. Manchester: Manchester University Press.

Weale Albert 2002. "Environmental Rules and Rule Making in the European Union", in Andrew Jordan (Ed.) *A Guide to EU Environmental Policy: Actors, Institutions and Processes*. pp. 198-213. London: Earthscan.

Weiss Carol H. 1998. "Have We Learned Anything New about the Use of Evaluation?". *American Journal of Evaluation* 19(1): 21-33.

Weiss Carol H. 2000. "Which Links in Which Theories Shall We Evaluate?", in Rogers Patricia J., Timothy A. Hacsi, Anthony Petrosino and Tracy A. Huebner (eds) *Program Theory in Evaluation: Challenges and Opportunities*. New Directions for Evaluation 87. pp. 35-45. San Francisco, CA: Jossey-Bass.

Wildavsky Aaron 1984. "Preface to the Third Edition: Implementation and Evaluation as Learning", in Jeffrey L. Pressman and Aaron Wildavsky (1984) *Implementation: How Great Expectations in Washington are Dashed in Oakland*. Third expanded edition. pp. xv–xviii. Berkeley, CA: University of California Press.

Working group for assessing the implementation of the Environmental Protection Act 2003. *Ympäristönsuojelulain täytäntöönpano: Arvio lain toimiivuudesta* [Implementation of the Environmental Protection Act: Assessment of Experiences]. Suomen ympäristö 661. Helsinki: Ministry of the Environment.

Yin Robert K. 1994. *Case Study Research: Design and Methods*. Second edition. Applied Social Research Methods Series Vol. 5. London: Sage Publications.

Zito Anthony 2002. "Task Expansion: A Theoretical Overview", in Andrew Jordan (Ed.) *A Guide to EU Environmental Policy: Actors, Institutions and Processes*. pp. 159-179. London: Earthscan.